MPRA

Munich Personal RePEc Archive

Forecasting 2016 US Presidential Elections Using Factor Analysis and Regression Model

Sinha, Pankaj and Srinivas, Sandeep and Paul, Anik and Chaudhari, Gunjan

Faculty of Management Studies, University of Delhi

25 July 2016

FORECASTING 2016 US PRESIDENTIAL ELECTIONS USING FACTOR ANALYSIS AND REGRESSION MODEL

Pankaj Sinha, Sandeep Srinivas, Anik Paul and Gunjan Chaudhari

Faculty of Management Studies, University of Delhi

ABSTRACT

The paper categorizes factors responsible for forecasting the outcome of U.S. presidential election 2016 using factor analysis, which groups the various economic and non-economic parameters based on the correlation among them. The major economic factor significant in 2016 US presidential election is the growth of the economy, and the 'anti-incumbency factor that signifies how long the incumbent party has been controlling the White House is found to be an important non-economic factor likely to play a dominant role in the election.

The dependent variables considered are the vote shares of the nominees of the incumbent and the non-incumbent majority party candidates. The forecast is calculated by running a regression of the significant factors, obtained through factor analysis technique, on the incumbent party vote share as well as on the non-incumbent party vote share.

The proposed models forecast the vote share of Democrat candidate Mrs. Hillary Clinton to be 45.59% with a standard error of ±2.32% and that of Republican candidate Mr. Donald Trump to be 39.51% with a standard error of ±3.87%. Hence, the models built in the paper signal a comfortable margin of victory for the Presidential nominee of the incumbent party, Hillary Clinton.

The study re-establishes the notion that the non-economic factors have a greater influence on the outcomes of election as compared to the economic factors, as some of the important economic factors such as inflation and unemployment rate failed to establish their significance.

INTRODUCTION

The United States Presidential election 2016 has drawn attention worldwide. The U.S. economy being strongly connected to most of the nations, the transitions in the country affect almost all the nations in the world. Because of this, forecasting the outcome of US Presidential elections is something that has been widely researched upon. Numerous statistical models have hence been brought forward in the past for this purpose. Some researchers employ methods that consider the economic variables while some use various non-economic variables as independent variables. Each method has its own merits and demerits. In this

paper, we explore the influence of various macroeconomic and non-economic variables on the Presidential elections forecasting model.

The Fair (1978, 2016) model forecasts the outcome of U.S. Presidential elections on the basis of economic parameters such as growth rate of real GDP per capita. The three economic variables used in this model are:

a) Growth rate of real per capita GDP in the first three-quarters of election year

b) Growth rate of the GDP deflator in the first 15 quarters of the president's administration

c) Number of quarters among the first 15 quarters of the president's administration in which the real rate of growth of the annual per capita GDP is higher than 3.2%.

Along with GDP, the unemployment rate has also been considered to be an important parameter in U.S. Presidential election. Several researchers have considered this as a significant parameter to gauge the dissatisfaction among the population. But Silver (2011) clarifies with examples that only minimal correlation exists between unemployment and the margin of victory whereas Jérôme and Jérôme (2011) model consider unemployment rate as the most significant parameter amongst all the economic factors.

Inflation is also considered as another major economic variable affecting U.S. elections. The fiscal model of U.S. presidential elections, developed by Cuzán, Heggen, and Bundrick (2016) and Cuzán (2000) use four of the six variables from Fair's model. In the fiscal model, the parameters considered include the growth rate of real GDP per capita and the type of fiscal policy (expansionary/contractionary) adopted by the incumbent party. From the models mentioned above, we can see inflation as one of the influential parameters in forecasting U.S. elections.

In past studies, we have observed the effect of macroeconomic variables like growth rate, unemployment, gold prices and inflation on the U.S. Presidential elections. But along with the economic parameters, non-economic parameters have also been found to play a vital role in the U.S. Presidential elections. Abramowitz (1988) model considers the "time for change" factor which takes into account the number of terms the incumbent party was in power. According to Abramowitz (1988) model, the time for change factor in 2016 is in favor of the Republican Party candidate (Mr. Donald Trump) as Democratic Party has been in power for two consecutive terms (Mr. Barack Obama). The present work labels this factor as the "anti-incumbency effect", owing to the fact that it helped the non-incumbent party in the elections.

Another crucial non-economic factor is whether the country was involved in any military interventions in the recent past. This parameter is prominent because it portrays the furtherance of United States' national interest along with the standing of the country in the world. War/Peace was used as a parameter by Yale economist Fair (1978, 2016), Lichtman and Keilis-Borok (1981) and in the "Bread and Peace" model by the political scientist Douglas Hibbs (2000, 2012).

The next non-economic factor is funds raised by the Presidential nominee in the overall campaign including the primaries. Weingert and Sebastian (2015) conclude that the funds from individual donors are a better predictor of winning than the total funds raised by the Presidential candidate. Presidential popularity is a key non-economic parameter in the model for forecasting the result of U.S. Presidential elections. Lewis-Beck and Rice (1982) included this factor based on Seigelman's (1979) research that indicated a correlation between the popularity rating of the reigning President and the upcoming election. Also, Seigelman (1979) provided a relationship between the popular vote share of the incumbent party and the Gallup rating as derived from the latest pre-election poll. Lewis-Beck et al (1982) proposed the use of Gallup ratings as acquired in the month of June of the election year. The state-by-state political economy model by Jérôme and Jérôme (2011) also considers the factor of incumbent President's credibility.

The June ratings are taken into account because they are sufficiently close enough to the election to provide a recent popularity rating of the incumbent President, and far enough from the elections to avoid what is usually termed as the 'electoral mood swings'. This relationship is clearly discussed by Lewis-Beck et al (1982). It is also important to note that though the Gallup ratings provide a popularity rating of the President that is based on extensive surveys, they certainly can't be considered exhaustive and there is, hence, an evident need for consideration of many other non-economic parameters for determining the outcome of an election.

The models developed by Tufte (1975, 1978) and Hibbs (1982) explain that the midterm elections in the Congress are seen as a referendum on the incumbent party. The two years of presidency is assessed by the voters by raising economic and non-economic issues. Along with all these parameters, non-economic factors such as demographics and scandal are also prominent. Demographic parameters considered include the percentage of vote share to the incumbent party by the white, male, female and youth (18-30 years of age).

Josh Katz (2016) mentions that out of the 50 states in the United States, 19 states strongly support Democratic party, whereas, 20 states strongly support the Republican party. The remaining states are termed as swing states; they can either be Democratic or Republican. The total number of electoral voters in U.S. is

538. The 19 Democratic states comprise for 228 electoral seats and the other 20 Republican states comprise for 155 electoral seats. The cities constituting higher share of electoral votes are in Democrats favour, for example California and New York encompass 55 and 29 electoral voters respectively[1].

To summarize our discussion on non-economic factors, anti-incumbency effect, male, white, youth, female, average Gallup rating, June Gallup rating, Gallup index, military intervention, funds raised, mid-term performance and scandal are the major non-economic variables which impact the U.S. Presidential elections. Sinha et al (2008, 2012) used Bayesian Prediction methods and factor analysis on economic and non-economic parameters for forecasting the outcome of U.S. Presidential election.

In this paper, we have integrated the economic and the non-economic variables and created a model, out of which only the significant factors are considered for forecasting the winner of U.S. Presidential election 2016.

SIGNIFICANCE OF THE VARIABLES CONSIDERED

This section deals with the significance of factors which have been grouped into two major categories of variables - Economic and Non-Economic.

Economic Variables:

The variables which have a direct bearing on the economy of a country have been clubbed as economic variables and the factors that do not have a direct impact on the economy have been clubbed as non-economic variables. Some factors like unemployment, GDP growth rate, inflation rate are indicators of the quality of work the incumbent President and the ruling party has done, while other factors such as gold prices indicate how U.S. has performed as an economy against the rest of the economies of the world. Anything related to the budget of United States of America points to the effective management of the economy by the present Government. Thus, the economic factors might have a definitive bearing on the election results.

The model deployed here can be divided into two separate steps. First, various variables were identified which could affect the election outcome. These variables were clubbed together into various factors using

---

[1]California and New York are strongly Democratic states

the Factor Analysis tool available in SPSS package. The next step entailed estimation of equations using the factors from the previous step to end up with an appropriate model for forecasting results of US Presidential elections. This was done for the incumbent party as well as for the non-incumbent party to get a clear idea about the outcome of the election. The model used to forecast the final vote share after incorporating the factors as independent variables was the Regression Analysis model. The following economic variables have been used in the model:

1. Unemployment Rate: Average annual rate of unemployment (percent) of labour force i.e. 16 years and over, as defined, by Bureau of Labour Statistics (2012a) is used.

2. Gold Prices: Inflation adjusted gold prices, yearly average taken from the website of the National Mining Association (US) is considered.

3. Growth Rate: The growth rate of the real per capita GDP in the average of the first three quarters of the corresponding election year (annual rate) is taken from Fair (2006).

4. Inflation: The annual CPI inflation rate averaged for the first six months of the election year is considered.

Non-Economic Variables

Demographic Factors

White Support: This factor denotes the ratio of the percentage of the white population supporting the incumbent party to the percentage of the white population supporting the non-incumbent party. The data for the past election years has been taken from the Gallup website. The data for the present election is taken from various online opinion polls including those on Gallup website. The same ratio for other races within the population were also considered, but were not found to significantly alter the vote share percentage of the incumbent party and were hence not used for the analysis. While building the model for the non-incumbent party, the reciprocal of the same ratio is used.

Male Support: This factor denotes the ratio of the percentage of male population supporting the incumbent party to the percentage of the male population supporting the non-incumbent party. The data for the past election years has been taken from the Gallup website. The data for the present election is taken from various online opinion polls, including the Gallup website. While building the model for the non-incumbent party, the reciprocal of the same ratio was used.

Youth Support: The percentage of youth population (18-30 years of age) supporting the incumbent (or non-incumbent party) is considered.

Female Support: This factor denotes the ratio of the percentage of the female population supporting the incumbent party to the percentage of the female population supporting the non-incumbent party. Again, the data for the past-election years has been taken from the Gallup website. The data for the present election is taken from various online opinion polls, including the Gallup website. While building the model for the non-incumbent party, the reciprocal of the same ratio was used.

Popularity Factors

Average Gallup Rating: This variable is a measure of the Presidential approval rating. The approval rating for the incumbent President is available in the form of a percentage value, which shows the percentage of the population approving of his/her work in the White House. Average Gallup Rating signifies the average of the Presidential approval rating throughout the tenure of the reign of the President.

Gallup Index: This variable again signifies the approval rating of the President, but classifies it into factors based on the average Gallup rating of the President. The same variable was used for predicting the outcome of the 2012 US presidential elections in Sinha et al (2012).

- For Average Gallup≤40: Index = 0
- 40<Average_Gallup<60: Index = 1
- Average_Gallup≥60: Index=2

June Gallup Rating: This variable signifies the approval rating of the President in the month of June of the election year. Since people are expected to vote based on the recent popularity of the President, June Gallup rating can be deemed to be of considerable significance in the determination of the vote share of the incumbent party.

Factors affecting popularity

Military Intervention: The factor shows the impact of military interventions during the tenure of the President on public perception of the incumbent party. This variable takes values as follows:

- +1, if military intervention improved the incumbent popularity
- -1, if military intervention decreased the incumbent popularity

- 0, if the intervention had no significant impact on the incumbent popularity

Mid-term performance:

The mid-term performance variable is the same as the one defined in Sinha et al (2012), for predicting the outcome of 2012 United States of America Presidential elections. The degree of fall in the approval rating of the incumbent President after the mid-term elections is found to be an indicator of the incumbent party's future in past studies. The popularity of the incumbent party almost always goes down after the mid-term elections, and is considered by many to be a referendum on the incumbent party. Mid-term variable is defined as:

For a particular year,

$$\text{Midterm} = \frac{\text{House}_{\text{Seats}} * \text{House}_{\text{Result}} + \text{Senate}_{\text{Seats}} * \text{Senate}_{\text{Result}}}{\text{House}_{\text{Seats}} + \text{Senate}_{\text{Seats}}}$$

House$_{\text{Seats}}$: Number of seats in the House taken by the Democrat and Republican Party representatives during the midterm election immediately before the corresponding final U.S. Presidential election

Senate$_{\text{Seats}}$: Number of seats in the Senate taken by the Democrat and Republican Party senators during the midterm election immediately before the corresponding final U.S. Presidential election

House$_{\text{Result}}$: This variable takes values from the set {-1, 0, 1} as given below:

- 1 if the incumbent party is in majority in the House after the midterm election immediately before the corresponding final U.S. Presidential election

- -1 if the incumbent party is in minority in the House after the midterm election immediately before the corresponding final U.S. Presidential election

- 0 otherwise,

Similarly, the variable **Senate$_{\text{Result}}$** takes values from the set {-1, 0, 1} as given below:

- 1 if the incumbent party holds majority in the Senate after the midterm election immediately before the corresponding final U.S. Presidential election

- -1if the incumbent party holds minority in the Senate after the midterm election immediately before the corresponding final U.S. Presidential election

- 0 otherwise

On an average, the number of seats in House is close to four times of the seats in Senate. Hence, the variable is tilted towards the values of **House_Seats**. Rather than being a flaw, it is considered as a better estimator of the image of the incumbent party and acts as a referendum. This is due to the fact that in midterm elections, all the seats of the House are considered for re-election, while in Senate, on an average; only 33% of the seats are being contested for re-election.

Scandal: Denotes the severity of the impact of a scandal that occurred during the term of the incumbent President. The scandals considered here may be political, personal or belonging to any other category. This variable takes values as follows:

- +1, if the scandal improved the incumbent popularity

- -1, if the scandal decreased the incumbent popularity

- 0, if the scandal had no significant impact on the incumbent popularity

Other factors

President Running: The variable shows if the incumbent President is a candidate for the ongoing election. A value of 1 is assigned to this variable, if the President is a candidate in the ongoing election and 0 otherwise.

Anti-incumbency effect: The factor takes into account the people's desire for a change in the ruling party. This factor has played a major role in the past in determining the outcomes of the elections. The factor takes a value 1 if the incumbent party has been in power for eight years or more (two or more terms), and 0 otherwise.

Fund raised: This factor is based on the notion that a Presidential candidate, who is unlikely to win, would not garner sufficient donations and funds. The difference in funds raised is significantly different in the forthcoming election and hence should not be ignored. The values of this variable have been obtained from the final campaign spending values for the past election years. To quantify this variable, the natural log of the ratio of spending (fund raised) of the incumbent party to that of the non-incumbent party is considered.

METHODOLOGY

Factor Analysis

Factor analysis is a statistical technique that has varied usage across a plethora of sectors. It is primarily a form of data reduction where the correlation between a certain set of independent or dependent variables is measured to find its combined effect on the dependent variable. One of the inherent problems in regression analysis is that with multi-collinearity in place, it is very difficult to estimate a suitable equation and the corresponding meaningful coefficients. Factor analysis helps us do away with this problem by finding the latent variables which are rather unobservable and their effect on the dependent variable. In case of factor analysis, an increase in the number of variables will not necessarily lead to an increase in the real vector space dimension, which is an added advantage over the normal regression analysis.

The fundamental objective of analyzing various variables on the basis of their grouped factors is to seek out for the inherent traits among the variables. The end result is the clubbing of all such uncorrelated variables to yield a minimum number of definite factors which aptly represent all related variables. These factors are thus a group of latent variables having fundamentally similar traits. There are several methods that are used to group the set of variables into multiple factors.

While finding the factors in the factor analysis process, the eigen values or the characteristic roots determine the variance among the variables captured by that particular factor. The factors which qualify the threshold eigen value ascribed by the user are the ones which are taken into consideration. The factors which have a low eigen value denote a lower contribution towards the variances in the variables captured by it and thus tend to be redundant. There are several rotation methods used while determining the factors from the given set of variables. Some of them maximise the variance among the variables captured by the factors while the others minimise the number of factors required to adequately capture the different variables.

The equations derived from factor analysis produce better coefficients in terms of econometric analysis. This is because of the amount of data in terms ofnumber of variables and variance that it takes into account while coming up with the factors and simultaneously compensating for errors and invariability at the same time.

DATA SOURCES

The Gallup ratings for the incumbent Presidents are available from the year 1948. The data has been considered from 1952-2016, because of the unavailability of data for some of the fields prior to that. The growth rate of GDP is taken from Fair (2006, 2008, and 2012). The inflation rate was taken from the website www.usinflationcalculator.com. The data used for inflation is the average of the first six months of the corresponding election year. Unemployment data is taken from the Bureau of Labor Statistics. The data on public debt (deficit/surplus) is obtained from the website of International Monetary Fund. The interest rate considered here has been taken from the Economic Data of Federal Reserve Bank of St Louis. The rate denotes the yield of the 3-month Treasury bill of the US Fed as on 1st of June of the election year.

Data for non-economic variables showing Presidential popularity such as average Gallup rating, Gallup index and June Gallup rating have been obtained from the official website of the Gallup Presidential poll. Data for other non-economic variables such as the male support, female support, white support, youth support have been taken from the official website of Gallup Presidential poll as well. The data on wars, scandals, military intervention and mid-term performance have been taken from Sinha et al (2012). The historical vote share data was taken from the website www.uselectionatlas.org. The data for the present year for the demographic variables like the male support, female support, youth support and white support have been taken from various opinion polls on the Gallup website, and the same have been employed for forecasting.

EMPIRICAL ANALYSIS OF MODELS

Incumbent Party Vote Share Analysis:

In the model developed, the anti-incumbency effect is considered as a different factor and rest of the variables clubbed into factors using factor analysis. The factor analysis tool in the SPSS package organized the variables into 5 factors – Factor1 comprises of male support, white support, female support, and growth rate of GDP. Factor2 consists of the average Gallup rating of the incumbent President, Gallup index and the Gallup rating of the incumbent President during the month of June of the election year. Factor3 comprises of a nominal variable to determine if the President is running or not, unemployment rate, military intervention and inflation rate. Factor4 contains the natural logarithm of the spending of incumbent party (funds raised for the current incumbent party), the natural logarithm of the gold price for one troy ounce (31.1 grams) of gold during the election years and youth support. Factor5 consists of the mid-term performance of the incumbent President and a variable that shows the change in popularity of the President due to scandals during the tenure of his incumbency.

Table 1: Rotated Component Matrix[a]

Factors	F1	F2	F3	F4	F5
			Component		
Male	.936				
White	.890				
Growth_of_economy	.772				
Female	.595	.534		.456	
Avg_Gallup_Rating		.907			
Gallup_Index		.890			
June_Gallup_Rating		.815			
President_running			.821		
Unemployment_rate			.750		
Military_intervention		.411	.628		
Inflation		-.491	.620		
Fund_raised				-.870	
log_gold				-.801	
Youth				.485	
Mid_term_performance					-.889
Scandal	-.507				.539

Extraction Method: Principal Component Analysis.

 Rotation Method: Varimax with Kaiser Normalization.

Table 2: Impact of combination of factors on vote share of incumbent party share using regression analysis

Sl. No	Factors	P value	R-squared
1	F2	0.6227	68.29
	F3	0.0858	
	F4	0.3074	
	F5	0.0527	
	ANTI-INCUMBENCY EFFECT	0.0041	
2	F1	0.0019	88.19
	F3	0.1848	
	F4	0.0669	
	F5	0.0312	
	ANTI-INCUMBENCY EFFECT	0.0037	
3	F1	0.0000	94.55
	F2	0.0005	
	F3	0.5660	
	F4	0.0272	
	ANTI-INCUMBENCY EFFECT	0.2778	
4	F1	0.0000	96.49
	F2	0.0013	
	F3	0.7406	
	F4	0.0101	
	F5	0.0527	
	ANTI-INCUMBENCY EFFECT	0.0354	
5	F1	0.0000	96.44
	F2	0.0003	
	F4	0.0059	
	F5	0.0359	
	ANTI-INCUMBENCY EFFECT	0.0076	

Non-Incumbent Party Vote Share Analysis:

The factor analysis tool in the SPSS package sorted the variables into 6 factors – Factor1 comprises of the average Gallup rating of the incumbent President, Gallup index and the Gallup rating of the incumbent President during the month of June of the election year. Factor2 consists of male support, white support and growth rate of GDP. Factor3 contains a nominal variable to determine if the President is running or not, unemployment rate, military intervention and inflation rate. Factor4 comprises of natural logarithm of the spending of incumbent party (funds raised for the current non-incumbent party), natural logarithm of the gold price for one troy ounce (31.1 grams) of gold during the election years. Factor5 consists of female support and youth support while Factor6 contains the mid-term performance of the incumbent President and a variable that shows the change in popularity of the President due to scandals during the tenure of his incumbency.

Table 3: Rotated Component Matrix

	Component					
	1	2	3	4	5	6
Avg_Gallup_Rating	.876					
Gallup_Index	.861					
June_Gallup_Rating	.780					
Male		.909				
White		.863				
Growth_of_economy		-.811				
President_running			.865			
Unemployment_rate			.800			
Inflation	-.424	.415	.621			
Military_intervention	.515		.516			
Fund_raised				.897		
log_gold				.880		
Youth					.960	
Female	-.541				.622	
Mid_term_performance						.908
Scandal		.478				-.496

Table 4: Impact of combination of factors on vote share of non-incumbent party share

Sl. No	Factors	P value	R-squared
1	F1	0.0242	89.56
	F2	0.0013	
	F3	0.5163	
	F4	0.6231	
	F5	0.0033	
	F6	0.9284	
	ANTI-INCUMBENCY EFFECT	0.6255	
2	F1	0.012	89.55
	F2	0.0005	
	F3	0.4821	
	F4	0.5823	
	F5	0.0013	
	ANTI-INCUMBENCY EFFECT	0.5656	
3	F1	0.0074	89.17
	F2	0.0003	
	F3	0.4627	
	F5	0.0004	
	ANTI-INCUMBENCY EFFECT	0.4454	
4	F1	0.0067	88.54
	F2	0.0002	
	F5	0.0002	
	ANTI-INCUMBENCY EFFECT	0.6523	
5	F1	0.0033	88.32
	F2	0.0001	
	F5	0.0001	

PROPOSED MODEL

The final model for determining the elections should satisfy the following criteria – achieve a high significance level for the coefficients of the factors formed by the combination of different variables, high value of coefficient of determination (R^2), acceptable levels of Root Mean Square (RMSE<1) and a Theil Statistic very close to zero (for a reliable prediction). Along with the above mentioned criteria for the final model, the factors calculated by SPSS 22 software should satisfy the criteria specified by the Kaiser-Meyer-Olkin test, which is KMO > 0.5.

Incumbent Party Vote Share Analysis:

The factor analysis was performed on a good number of economic and non-economic variables. The anti-incumbency effect factor was considered separately as it was found to be significant on its own. The final model (Model 5 in Table 2) proposes to consider Factor1, Factor2, Factor4, Factor5 and the anti-incumbency effect. Factor3 consisting of the variables mentioned before, was found to be insignificant and hence was not considered for the final model building. Hence, the proposed model is given as:

$$Y = c + \alpha_1\, \text{Factor1} + \alpha_2\, \text{Factor2} + \alpha_4\, \text{Factor4} + \alpha_5\, \text{Factor5} + \alpha_6\, (\text{anti-incumbency effect}) + \varepsilon$$

The dependent variable Y in this case, represents the vote share of the incumbent party. This model can be used to forecast the vote share of the incumbent party for the year 2016, using the regression model built.

Model used for forecasting vote share of the incumbent party

$$\text{VOTE} = c + \alpha_1\, \text{Factor1} + \alpha_2\, \text{Factor2} + \alpha_4\, \text{Factor4} + \alpha_5\, \text{Factor5} + \alpha_6\, (\text{anti-incumbency effect}) + \varepsilon$$

$$\text{VOTE} = 53.269 + 3.922\, \text{Factor1} + 2.014\, \text{Factor2} + 1.374\, \text{Factor4} + 0.921\, \text{Factor5} - 3.088\, (\text{anti-incumbency effect})$$

Dependent Variable: INCUMBENT_VOTE_SHARE

Method: Least Squares

Sample: 1952 2012

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	P-Value
C	53.26910	0.557090	95.62029	0.0000
F1	3.921713	0.371093	10.56801	0.0000
F2	2.014146	0.367770	5.476643	0.0003
F4	1.373917	0.394149	3.485783	0.0059
F5	0.921256	0.380306	2.422406	0.0359
ANTI-INCUMBENCY EFFECT	-3.088278	0.927264	-3.330527	0.0076

R-squared	0.964482	Mean dependent var	52.01213
Adjusted R-squared	0.946723	S.D. dependent var	5.596522
S.E. of regression	1.291780	Akaike info criterion	3.629916
Sum squared resid	16.68696	Schwarz criterion	3.919637
Log likelihood	-23.03933	Hannan-Quinn criter.	3.644752
F-statistic	54.30935	Durbin-Watson stat	2.426019
Prob(F-statistic)	0.000001		

The coefficient of determination (R^2) value obtained for the model was found to be 96.45%, which is high enough for a good prediction of the vote share of the incumbent party in the 2016 Presidential elections. The Theil statistic for the forecast was found to be 0.0097, which is good enough for an accurate forecast. Using data from the same sources as mentioned for the economic variables and Gallup ratings and obtaining the data for some non-economic variables like male support, female support, youth support and white support from various online polls, the model was employed for prediction of the vote share of the domestic candidate. The youth support value for 2016 was taken as 31%. The male support, female support and white support ratios were taken to be 0.75, 0.79 and 0.89 respectively. The final forecast for the vote share of incumbent party obtained was 45.586%, with a standard error of ± 2.32%.

Non-Incumbent Party Vote Share Analysis:

The forecasted vote share of 45.586% for the candidate of the incumbent party is not sufficient to forecast the final outcome of the election, i.e. the next President of the United States. Hence, a similar analysis was carried out to forecast the vote share of the Presidential candidate of the non-incumbent party.

Model used for forecasting vote share of the non-incumbent party

$VOTE = c + \alpha_1 \text{ Factor1} + \alpha_2 \text{ Factor2} + \alpha_5 \text{ Factor5} + \varepsilon$

$VOTE = 45.453 - 1.829\text{Factor1} + 2.959 \text{ Factor2} + 3.602\text{Factor5}$

Dependent Variable: NON_INCUMBENT_VOTE_SHARE

Method: Least Squares

Sample: 1952 2012

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	P-Value
C	45.45282	0.507872	89.49654	0.0000
F1	-1.829641	0.500880	-3.652849	0.0033
F2	2.959331	0.508300	5.822017	0.0001
F5	3.602167	0.606188	5.942326	0.0001

R-squared	0.883182	Mean dependent var	45.82437
Adjusted R-squared	0.853978	S.D. dependent var	5.239949
S.E. of regression	2.002333	Akaike info criterion	4.438821
Sum squared resid	48.11203	Schwarz criterion	4.631968
Log likelihood	-31.51056	Hannan-Quinn criter.	4.448711
F-statistic	30.24141	Durbin-Watson stat	1.676520
Prob(F-statistic)	0.000007		

The coefficient of determination (R^2) for the model was found to be 88.32%, which is high enough for a good prediction of the 2016 vote share of the non-incumbent party. The Theil statistic for the forecast was found to be 0.0188, which is good enough for an accurate forecast. Using data from the same sources as mentioned for the economic variables and Gallup ratings and obtaining the data for some non-economic variables like male support, female support, youth support and white support from various online polls, the model was employed for prediction of the vote share of the domestic candidate. The final forecast for the vote share that was obtained was 39.508%, with a standard error of ±3.87%.

CONCLUSION

This paper uses the techniques of factor analysis and regression to build models capable of forecasting the outcome of the US Presidential Election of 2016. According to the analysis performed, multiple factors that included a mixture of economic and non-economic parameters were found to be significant for predicting the outcome of the forthcoming election.

The models developed in the paper forecast the vote share of the Democratic Party's candidate Hillary Clinton to be 45.59% with a standard error of ±2.32% and the vote share of the Republican Party's candidate Donald Trump to be 39.51% with a standard error of ±3.87%. From the vote share percentages predicted, we can say that Hillary Clinton is going to be the probable winner in this year's race to the White House, by a comfortable margin. These vote percentages also signal a rise in the percentage of votes received by the minority party nominees including Gary Johnson and Jill Stein, totaling to about 15%. Again, this paper re-establishes the notion that the non-economic factors have a greater influence on the outcomes of US Presidential elections as compared to the economic factors, as some of the important economic factors such as inflation and unemployment rate were found to be insignificant.

REFERENCES:

Abramowitz A. I. (1988). An Improved Model for Predicting the Outcomes of Presidential Elections. *PS: Political Science and Politics*, 21 4, 843-847

Bureau of Labor Statistics. (2012a). How the Government Measures Unemployment retrieved from: http://www.bls.gov/cps/cps_htgm.htm#unemployed

Bureau of Labor Statistics. (2012b). *Where can I find the unemployment rate for previous years?* retrieved from: http://www.bls.gov/cps/prev_yrs.htm/

Bureau of Economic Analysis. (2012). Table 3.12. Government Social Benefits. retrieved from: http://www.bea.gov/national/index.htm#gdp

Cuzán, A. G., Heggen R.J., & Bundrick,C.M. (2000). Fiscal policy, economic conditions, and terms in office: simulating presidential election outcomes. *In Proceedings of the World Congress of the Systems Sciences and ISSS International Society for the Systems Sciences*, 44th Annual Meeting, July 16–20, Toronto, Canada.

Cuzán, A. G. (2016). Fiscal model forecast for the 2016 U.S. presidential election. retrieved from: http://pollyvote.com/wp-content/uploads/2016/08/FISCAL-MODEL-FORECAST-FOR-2016-AMERICAN-PRESIDENTIAL-ELECTION.pdf

Hibbs D. A. (2000). Bread and Peace voting in U.S. presidential elections. *Public Choice, 104,* 149–180.

Hibbs D. A. (2012). Obama's Re-election Prospects Under 'Bread and Peace' Voting in the 2012 US Presidential Election.retrieved from: http://www.douglashibbs.com/HibbsArticles/HIBBS_OBAMA-REELECT-31July2012r1.pdf

Fair, R. C. (1978). The effect of economic events on votes for president. *Review of Economics and Statistics, 60*, 159-173.

Fair, R. C. (2002). Predicting Presidential Elections and Other Things. *Stanford: Stanford University Press.*

Fair, R.C. (2006). *The Effect of Economic Events on Votes for President: 2004 Update*. retrieved from: http://fairmodel.econ.yale.edu/RAYFAIR/PDF/2006CHTM.HTM

Fair, R.C. (2008). *2008 Post Mortem*. retrieved from: http://fairmodel.econ.yale.edu/vote2008/index2.htm

Fair, R. C. (2012). *Vote-Share Equations: November 2010 Update.* retrieved from: http://fairmodel.econ.yale.edu/vote2012/index2.htm

Fair, R.C. (2014). Presidential and Congressional Vote-Share Equations: November 2014 Update retrieved from: https://fairmodel.econ.yale.edu/vote2016/index2.html

Gallup Presidential Poll. (2016). *Presidential Job Approval Center*. retrieved from: http://www.gallup.com/poll/124922/presidential-approval-center.aspx

International Monetary Fund. (2010). *Historical Public Debt Databas*e. retrieved from: http://www.imf.org/external/pubs/ft/wp/2010/data/wp10245.zip

InflationData.com. (2016). *Historical Crude Oil Prices (Table).* retrieved from: http://inflationdata.com/inflation/Inflation_Rate/Historical_Oil_Prices_Table.asp

Jérôme B. & Jérôme V. (2011). Forecasting the 2012 U.S. Presidential Election: What Can We Learn from a State Level Political Economy Model. *In Proceedings of the APSA Annual meeting Seattle*, September 1-4 2011.

Katz J.(2016) Who will be President? Retrieved from: http://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html

Lichtman, A. J., and Keilis-Borok, V. I. (1981). "Pattern Recognition Applied to Presidential Elections in the United States, 1860-1980: Role of Integral Social, Economic and Political Traits," *Proceedings of the National Academy of Science*, Vol. 78, No. 11, pp. 7230-7234

Lichtman, A. J. (2005). The Keys to the White House. *Lanham, MD: Lexington Books*.

Lichtman, A. J. (2008). The keys to the white house: An index forecast for 2008. *International Journal of Forecasting, 24*, 301–309.

Lewis-Beck, M. S. & Rice, T. W. (1982).Presidential Popularity and Presidential Vote. *The Public Opinion Quarterly, 46 4,* 534-537.

Office of the Clerk. (2010). *Election Statistics*. retrieved from: http://artandhistory.house.gov/house_history/electionInfo/index.aspx

Sigelman, L. (1979). Presidential popularity and presidential elections. *Public Opinion Quarterly, 43,* 532-34.

Silver. N. (2011). On the Maddeningly Inexact Relationship Between Unemployment and Re-Election. retrieved from: http://fivethirtyeight.blogs.nytimes.com/2011/06/02/on-themaddeningly-inexact-relationship-between-unemployment-and-re-election/

Sinha P. and Bansal. A. K. (2008). Hierarchical Bayes Prediction for the 2008 US Presidential Election. *The Journal of Prediction Markets, 2,* 47-60.

Sinha P., Sharma A. & Singh H. (2012). Prediction for the 2012 United States Presidential election using multiple regression model, *The Journal of Prediction Markets, 6 2, 78-98.*

Tufte, E. R. (1975). Determinants of the Outcomes of Midterm Congressional Elections. *American Political Science Review*, 69, 812-26.

The White House. (2012). *Table 1.2—Summary of Receipts, Outlays, And Surpluses Or Deficits (–) As Percentages Of GDP: 1930–2017*. retrieved from: http://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist.pdf

United States National Mining Association. (2011). *Historical Gold Prices- 1833 to Present.* Retrieved from: http://www.nma.org/pdf/gold/his_gold_prices.pdf

Weingert & Sebastian(2015). Predicting primary election results through network analysis of donor relationships retrieved from: https://web.stanford.edu/class/cs224w/projects_2015/Predicting_primary_election_results_through_network _analysis_of_donor_relationships.pdf

APPENDIX

Table 1: Scandals during Presidential Terms and the Corresponding Ratings

Election Year	Incumbent President	Scandals	Scandal
1948	Franklin D. Roosevelt	- Budget cuts for the military - Recognition of Israel - Taft- Harley Act: Reducing the power of the labor unions	1
	Harry S. Truman	None	
1952	Harry S. Truman	- Continuous accusations of spies in the USGovt. - Foreign policies: Korean war, Indo China war - White house renovations - Steel and coalstrikes - Corruption charges	1
1956	Dwight D.	None	0
1960	Dwight D. Eisenhower	- U-2 Spy Plane Incident - Senator Joseph R. McCarthy Controversy - Little Rock School Racial Issues	1
1964	John F. Kennedy	Extra marital relationships	0
	Lyndon B. Johnson	None	
1968	Lyndon B. Johnson	- Vietnam war - Urban riots - Phone Tapping	1
1972	Richard Nixon	Nixon shock	0
1976	Richard Nixon	Watergate Scandal	2
	Gerald Ford	Nixon Pardon	
1980	Jimmy Carter	- Iran hostage crisis 1979 energy crisis - Boycott of the Moscow Olympics	1
1984	Ronald Reagan	Tax cuts and budget proposals to expand military spending	0
1988	Ronald Reagan	- Iran-Contra affair - Multiple corruption charges against high ranking officials	1
1992	George H. W. Bush	- Renegation on election promise of no new taxes "VomitingI ncident"	1
1996	Bill Clinton	- Firing of White House staff "Don't ask, don't tell "policy	1
2000	Bill Clinton	Lewinsky Scandal	2
2004	George W. Bush	Poor handling of Katrina Hurricane-None	0
2008	George W. Bush	- Midterm dismissal of 7 US attorneys - Guantanamo Bay Controversy and torture	1
2012	Barack Obama	None	0
2016	Barack Obama	None	0

Table 2: Military Interventions during Presidential Terms and the Corresponding Ratings

Election Year	Incumbent President	Military Interventions	War Ratin
1948	Franklin D. Roosevelt	• World War2	1
	Harry S. Truman	• Hiroshima/Nagasaki	
1952	Harry S. Truman	• Korean War	-1
1956	Dwight D. Eisenhower	• Ended Korean War	1
1960	Dwight D. Eisenhower	• None	0
1964	John F. Kennedy	• Bay of Pigs • Cuban Missile crisis • Vietnam	-1
	Lyndon B. Johnson	• Vietnam	
1968	Lyndon B. Johnson	• Vietnam • Israel	-1
1972	Richard Nixon	• Vietnam	-1
1976	Richard Nixon	• Vietnam	1
	Gerald Ford	• Vietnam(end)	
1980	Jimmy Carter	• None	0
1984	Ronald Reagan	• Cold War	0
1988	Ronald Reagan	• Cold War	0
1992	George H. W. Bush	• Panama • Gulf War • Somalia	-1
1996	Bill Clinton	• Somalia • Bosnia	0
2000	Bill Clinton	• Serbians(Yugoslavia)	0
2004	George W. Bush	• Afghanistan • Iraq	1
2008	George W. Bush	• Afghanistan • Iraq	-1
2012	Barack Obama	• Ended Iraq war • Increased presence in Afghanistan • Military Intervention in Libya	1
2016	Barack Obama	• War against ISIS • Ground troops to Iraq	-1

Table 3: Gallup Ratings

Election Year	Incumbent President	Period of Gallup Measurement	Rating	June Gallup Rating	Average Gallup Rating	Gallup Index
1948	Harry S. Truman	May 27-June1	39	39.5	55.6	1
		June 17-23	40			
1952	Harry S. Truman	May 29-June 3	31	31.5	36.5	0
		June 14-19	32			
1956	Dwight D. Eisenhower	May 30-June 4	71	72	69.6	2
		June 14-19	73			
1960	Dwight D. Eisenhower	June 15-20	61	59	60.5	2
		June 29-July 4	57			
1964	Lyndon B. Johnson	June 3-8	74	74	74.2	2
		June 10-15	74			
		June 24-29	74			
1968	Lyndon B. Johnson	June 12-17	42	41	50.3	1
		June 25-30	40			
1972	Richard Nixon	June 15-18	59	57.5	55.8	1
		June 22-25	56			
1976	Gerald Ford	June 10-13	45	45	47.2	1
1980	Jimmy Carter	May 29-June 1	38	33.6	45.5	1
		June 12-15	32			
		June 26-29	31			
1984	Ronald Reagan	June 5-7	55	54	50.3	1
		June 21-24	54			
		June 28-July 1	53			
1988	Ronald Reagan	June 9-12	51	50	55.3	1
		June 23-26	48			
		June 30-Jul 6	51			
1992	George H. W. Bush	June 3-6	37	37.3	60.9	2
		June 11-13	37			
		June 25-29	38			
1996	Bill Clinton	June 17-18	58	55	49.6	1
		June 26-29	52			
2000	Bill Clinton	June 5-6	60	57.5	60.6	2
		June 21-24	55			
2004	George W. Bush	June 2-5	49	48.5	62.2	2
		June 20-22	48			
2008	George W. Bush	June 8-11	30	29	36.5	0
		June 14-18	28			
2012	Barack Obama	May 27-June 2	46	46.4	49.0	1
		June 3-9	47			
		June 10-16	46			
		June 17-23	46			

		June 24-30	47			
2016	Barack Obama	Jun 6-12	53	51.6	48.0	1
		May 30-Jun 5	51			
		Jun 13-19	53			
		Jun 20-26	50			
		Jun 27-Jul 3	51			

Source: Gallup Presidential Poll (2016)

Table 4: Midterm Elections Results (1944-2014)

Year	Incumbent Party	Midterm Election Year	House Seats		HouseResult	Senate Seats		SenateResult	Midterm Values
			Democratic	Republican		Democratic	Republican		
1948	Democratic	1944	243	190	-1	57	38	-1	-1.00
		1946	188	246		45	51		
1952	Democratic	1948	263	171	1	54	42	1	1.00
		1950	234	199		48	47		
1956	Republican	1952	213	221	-1	46	48	-1	-1.00
		1954	232	203		48	47		
1960	Republican	1956	234	201	-1	49	47	-1	-1
		1958	283	153		64	34		
1964	Democrat	1960	262	175	1	64	36	1	1.00
		1962	258	176		67	33		
1968	Democrat	1964	295	140	1	68	32	1	1.00
		1966	248	187		64	36		
1972	Republican	1968	243	192	-1	58	42	-1	-1.00
		1970	255	180		54	44		
1976	Republican	1972	242	192	-1	56	42	-1	-1.00
		1974	291	144		61	37		
1980	Democrat	1976	292	143	1	61	38	1	1.00
		1978	277	158		58	41		
1984	Republican	1980	242	192	-1	46	53	1	-0.63
		1982	269	166		46	54		
1988	Republican	1984	253	182	-1	47	53	-1	-0.63
		1986	258	177		55	45		
1992	Republican	1988	260	175	-1	55	45	-1	-1.00
		1990	267	167		56	44		
1996	Democrat	1992	258	176	-1	57	43	-1	-1.00
		1994	204	230		48	52		
2000	Democrat	1996	207	226	-1	45	55	-1	-1.00
		1998	211	223		45	55		
2004	Republican	2000	212	221	1	50	50	1	1.00
		2002	204	229		48	51		
2008	Republican	2004	202	232	-1	44	55	0	-0.82
		2006	233	202		49	49		
2012	Democrat	2008	256	178	-1	55	41	1	-0.63
		2010	193	242		51	47		
2016	Democrat	2012	200	234	-1	53	45	1	-0.63
		2014	188	247		44	54		

Source: Office of the Clerk (2010)

Table 5a: Macroeconomic Variables

Year	Growth of economy (%)[a]	Inflation (%)[b]	Unemployment rate (%)[c]	Gold_prices ($/ounce)[d]
1952	0.691	2.50	3	34.6
1956	-1.451	0.82	4.1	34.99
1960	0.377	1.58	5.5	35.27
1964	5.109	1.40	5.2	35.1
1968	5.043	3.92	3.6	39.31
1972	5.914	3.28	5.6	58.42
1976	3.751	6.22	7.7	124.74
1980	-3.597	14.40	7.1	615
1984	5.44	4.43	7.5	361
1988	2.178	3.93	5.5	437
1992	2.662	2.98	7.5	343.82
1996	3.121	2.80	5.4	387.81
2000	1.219	3.28	4	279.11
2004	2.69	2.33	5.5	409.72
2008	0.22	4.23	5.8	871.96
2012	1.62	2.37	8.075	1668.98
2016	1.1	1.07	4.9	1160.6

a: Fair (2006, 2008, 2012) b: usinflationcalculator.com, c: Bureau of Labor Statistics (2012b),

d: United States National Mining Association. (2011).

Table 5b: Non-economic Variables

Year	Anti-Incumbency effect	June Gallup Rating[a]	Avg Gallup Rating[b]	Gallup Index[c]	President running	Fund raised[d]
1952	1	31.5	36.5	0	0	1.61
1956	0	72	69.6	2	1	2.93
1960	1	59	60.5	2	0	2.31
1964	0	74	74.2	2	0	2.17
1968	1	41	50.3	1	0	2.45
1972	0	57.5	55.8	1	1	4.12
1976	1	45	47.2	1	1	3.58
1980	0	33.6	45.5	1	1	3.89
1984	0	54	50.3	1	1	4.21
1988	1	50	55.3	1	0	4.38
1992	1	37.3	60.9	2	1	4.58
1996	0	55	49.6	1	1	4.75
2000	1	57.5	60.6	1	0	4.79
2004	0	48.5	62.2	2	1	5.87
2008	1	29	36.5	0	0	8.48
2012	0	46.4	49	1	1	6.6
2016	1	51.6	48	1	0	5.39

a, b, c: Gallup Website, d: The History of Campaign Spending (Metrocosm website)

Table 5c: Non-economic Variables

Year	Male_Ratio[a]	Female_Ratio[b]	White_Ratio[c]	Youth[d]
1952	0.887	0.724	0.750	51
1956	1.222	1.564	1.439	57
1960	0.923	1.041	1.041	45.45
1964	1.500	1.632	1.439	64
1968	0.953	1.047	0.809	55.29
1972	1.703	1.632	2.125	52
1976	0.849	1.063	1.130	45.92
1980	0.717	0.898	0.643	53.41
1984	1.778	1.222	1.941	60
1988	1.273	1.083	1.439	63
1992	0.902	0.826	1.051	48.052
1996	1.138	1.385	1.022	64.29
2000	0.865	1.178	0.782	50
2004	1.273	1.083	1.273	40
2008	1.000	0.754	1.222	39
2012	0.887	0.754	0.786	62
2016	N/A	N/A	N/A	N/A

a, b, c, d: Gallup website(gallup.com)
Note: For Non-Incumbent vote share forecasting, reciprocals of a, b and c are used.

Table 6: Vote Share – Incumbent and Non-Incumbent

Year	Incumbent Party Vote Share[a]	Non-Incumbent Vote Share[b] (Major Opposition)
1952	44.33	55.18
1956	57.37	41.97
1960	49.55	49.72
1964	61.05	38.47
1968	42.72	43.42
1972	60.67	37.52
1976	48.01	50.08
1980	41.01	50.75
1984	58.77	40.56
1988	53.37	45.65
1992	37.45	43.01
1996	49.23	40.72
2000	48.38	47.87
2004	50.73	48.26
2008	45.6	52.86
2012	51.01	47.15

a, b : uselectionatlas.org