The Evolution of an Election Forecasting Model over 20 Years

Hans W. Ittmann, Institute for Transport and Logistics Studies (ITLS) Africa
University of Johannesburg, South Africa

Keywords: Clustering, Forecasting, Elections, Big Data

Extended Abstract:

1. Introduction

In 1999 a model was developed to predict the final elections outcome in South Africa based on early results as they are made available. The same basic model has been used since then for each of the subsequent national and local elections, which is held every five years respectively, in the country up to the last local elections in 2016. It is not often that operations researchers get the opportunity to remain involved with one specific model application as it is being used over such an extended period. It is thus interesting to retrospectively look at how the model performed over this period and how the use thereof has evolved over time. The political environment has not remained stable in South Africa during this period while both the model team and technology undergone changes as well.

In general the model has performed exceptional well over time. The model remained resilient in coping with changes in the political sphere such as the introduction of new political parties. However, in this regards during more recent elections there have been challenges identified around some of the assumptions made when the model was developed initially. Another exciting development was adapting the model for use in the 2016 Presidential election in the United States.

The aim of this paper is to present the model briefly, highlighting the use and performance of this election forecasting model. In addition challenges that were encountered along the way will be discussed as well as the adaption of the model for the US presidential election and its performance. Furthermore brief mention will be made of communicating and disseminating the election forecasting predictions of results to the media on radio, TV and electronically.

2. The Problem

Elections generate wide interest in a country and there is typically much speculation before an election in the media as to the possible outcome. There are many types of elections forecast. Before an election there are numerous market surveys conducted as well as electoral polls. All of these endeavours are aimed at giving an indication of what the thinking of voters are before an election. Then there are different scientific forecasting models that use historic election results with inputs from surveys to try to predict the final outcome of an election before the election. Some countries use exit polls, obtained from voters as they leave the voting stations, to predict the outcome after the voters have voted. Then there are election night forecasts, which endeavour to forecast the ultimate result before the final outcome is known, using early results as they are released.

South Africa introduced a system of proportional representation in 1994 for electing representatives for the National Assembly and for each of the nine Provincial Assemblies. In this type of system the

voters cast their vote for a political party and the number of seats gained by each party is proportional to the number of votes received by each party.

The forecasting model that was developed to predict the outcome of the South African elections is an election night forecast. As the first results from voting districts across the country are released, the model can be used for predicting the election results.

3. Model Formulation

The most appropriate way to model the voting population was to formulate a cluster model. The aim is to divide the population, or electorate, into groups with similar voting behaviour. These clusters are determined before the elections and are then used during the elections to extrapolate partial results to the whole cluster and thus the whole electorate. The fuzzy clustering approach of Bezdek (1980) is the clustering methodology that was implemented and has been used since the start.

The full formulation of the clustering method is given in Greben *et al*. (2005) and Greben *et al*. (2006). A suitable objective function is minimized, optimizing the positions of the cluster centres to ensure that the sum of distances squared between the cluster centres and the cluster members is minimal. It is important to point out that fuzzy clustering is used which implies each element has a distributional, rather than a discrete, membership of the clusters. The distinct advantage of this is that the distributional membership allows one to make predictions for all clusters, as soon as the first result is available.

In addition to the above the model is characterized by two major assumptions. Firstly, it is assumed that voting behaviour has a statistical pattern that can be modelled by clustering the voting districts (VDs), of which there are about 22 000 in total in South Africa, into segments with similar voting behaviour. Secondly, the order in which the voting results are declared is important. The model is designed to adjust for moderate non-random patterns in this order but will not operate well under extreme political or spatial bias.

4. Forecasting results

The forecasting results are briefly presented for 4 of the elections prior to 2014. For 2004 and 2009, the predictions for the national vote are presented in Table 1 while for the local, or municipal, elections in 2006 and 2011 the predictions are presented in Table 2 (in the case of local elections these predictions are for the PR (Proportional Representation) vote aggregated to a national level). The results in both Tables are only given for the 3 or 4 largest parties as they attract the most interest. The forecasts for the smaller parties are initially not very accurate and there is little interest for them.

From all four examples given here it is clear that the predictions were fairly accurate especially given the fact that in all cases the percentage votes counted were still low. The model performs better and the predictions become more stable and accurate as more VDs are counted, with the predictions converging to the final result once all VDs are declared. The results in the 2 Tables below, however, illustrate the accuracy of the model at an early stage of the vote counting process. The model behaved well up to the 2011 election.

From a political point of view there is much to be discussed given these predictions and in depth analysis can be performed using these. Some of this the team, or at least one or two of the team members, need to be able to contribute towards during such interactions in the media.

Party	2004				2009			
	Actual at 4.7% votes counted	Predicted at 4.7% votes counted	Deviance of predicted from final at 4.7% votes	Final Result	Actual at 5.0% votes counted	Predicted at 5.0% votes counted	Deviance of predicted from final at 5.0% votes	Final Result
ANC	60.88	69.76	0.08	69.68	61.19	65.59	0.32	65.91
DA	21.99	14.08	1.71	12.37	21.18	17.32	0.68	16.64
IFP	3.26	5.69	1.28	6.97	2.77	4.15	0.40	4.55
COPE					7.25	6.83	0.59	7.42

Table 1: Comparison of forecasts and actuals for 2004 and 2009 national elections.

Party	2006				2011			
	Actual at 6.8% votes counted	Predicted at 6.8% votes counted	Deviance of predicted from final at 6.8% votes	Final Result	Actual at 4.9% votes counted	Predicted at 4.9% votes counted	Deviance of predicted from final at 4.9% votes	Final Result
ANC	61.18	66.89	1.22	65.67	56.89	62.19	0.74	62.93
DA	24.32	17.35	1.03	16.32	32.28	25.94	1.86	24.08
IFP	2.82	6.28	1.28	7.56	1.58	3.02	0.54	3.56
COPE					3.28	2.95	0.73	2.22

Table 2: Comparison of forecasts and actuals for the 2006 and 2011 municipal elections.

5. New party Entrants

One of the first challenges faced by the model was the introduction of a new political party, COPE, in 2009 as shown in Table 1. From the good prediction, with only 5% of the votes counted, for this new entrant in that election (year 2009) it is clear the model behaved well and was clearly very resilient in coping with a new political party. In the local elections in 2011 where COPE was also contesting the local elections as a new entrant the same happened, thus illustrating again how resilient the model was.

In the 2014 elections, another new party emerged, namely the EFF (see Table 3 below). On the strength of the 2009 election prediction, with COPE as a new entrant, the expectation was that the model should cope well with this new entrant namely the EFF. This was not the case. With only 5% of the votes counted, the prediction for the DA and IFP were very close to the final outcome but this

was not the case for the ANC and the EFF. When 40.2% of the votes were counted the ANC was still 1.15% too high and the prediction for the EFF was still 1.56% too low. This seemed to indicate the model produced predictions that were less accurate than one would expect from the model.

Party	2014						
	Actual at 4.7% votes counted	Predicted at 4.7% votes counted	Absolute deviance between predicted and final at 4.7% votes	Actual at 40.2% votes counted	Predicted at 40.2% votes counted	Absolute deviance between predicted and final at 40.2% votes	Final result
ANC	53.85	63.83	1.68	61.74	63.30	1.15	62.15
DA	33.63	22.18	0.05	23.58	21.89	0.34	22.23
IFP	1.66	2.14	0.26	2.37	2.37	0.03	2.40
EFF	3.31	3.96	2.39	4.62	4.79	1.56	6.35

Table 3: Comparison of actual and predicted to final results for the 2014 national elections at roughly 5% and 40% of votes counted.

A thorough post-election analysis was conducted and this indicated that there were assumptions that did not hold (Ittmann *et al*. 2017).

6. Assumptions

An in depth analysis was conducted into the cause of this discrepancy that was observed in the 2014 national elections. It was possible to attributed this to two of the assumptions made in formulating the model. The clustering could have been more accurately determined beforehand and this would have made a difference. The poor performance of the ANC and EFF predictions as shown in Table 3 was the result of poorly constructed clusters due to the 2009 national results being a bad predictor of homogeneity in 2014 voter behaviour. This together with the fact that there was a delay in releasing the results of voting districts in the province (Gauteng) where the EFF had it biggest support had a direct negative impact on the accuracy of the predictions produced by the forecasting model (Ittmann *et al.*2017)

7. Forecasting a Foreign Election

There has always been the urge to see whether the model can be used in an election of another country. This opportunity arose in 2016 with the US presidential election. The US election system and process needed to be studied and understood. In addition it was essential to adapt and modify the model to ensure it is compatible with the US presidential election. In the US the number counties (where counties are the equivalent to VDs) are much less, dramatically less, than in South Africa and this had to be catered for in the model. The team needed to understand how the data representing the election in the USA looks like and where the election results are published online. A fair amount of effort was spent on ensuring the model was correctly adapted for use in the US presidential election. The graph below (Figure 1) shows that after 8.7% of the US counties were

counted and the results released, the model was able to correctly predict that Donald Trump would win the election (see Figure 1) (Duleni-Tlhone *et al*. 2018).

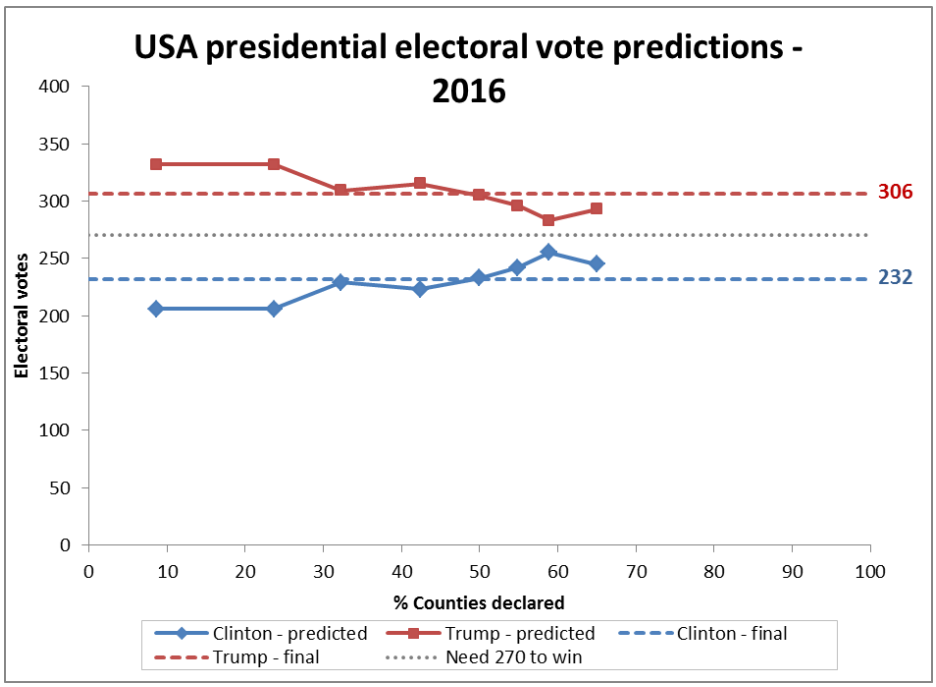

Figure 1: Final and predicted electoral vote counts for the 2016 US presidential elections at various percentages of counties declared

8. Communicating Election results

The election prediction results and the interpretation of these results need to be communicated through the media, both radio and TV, throughout the election process. There are previews before the elections and people want to know how the model works. When the results are made available a representative of the team needs to be available to appear "live" on radio programs as well as TV programs dedicated to the election. You need a person with insight into politics as well as understanding the way the model works and how to explain, not just the model, but also the results and assist in interpreting them.

9. Diverse Issues

Diverse issues that the team had to deal with which includes constant changing of team member composition. There is the natural coming and going of people while many of the original people have retired. The effect is that new people had to be brought into the team on a regular basis. They need to learn the model, learn the environment, etc. The technology used is also changing continuously. The programming language used initially was becoming outdated while the author of this software retired. This implied that the entire package had to be rewritten and tested before it could be used again. Printouts were used to communication the results to those interacting with the media. This is time consuming and new forecasts could not be updated immediately. The App, WhatsAppo, was used in the 2016 local elections to send updates of election predictions to the entire team and also those from the media. In this way the predictions as they became available were available almost immediately to all interested parties.

10. Conclusion

It has been a fascinating journey to be involved in this exciting project. The timeline of the project being "live" is very short lived but during the period when the first results are released till the ultimate result is known, the predictions obtained through use of the model remains of huge interest, till it becomes obvious what the final outcome of the election will be.

References

1. Bezdek JC, 1980, A convergence theorem for the fuzzy iso-data clustering algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(1), pp. 1-8.
2. Greben JM, Elphinstone C, Holloway J, De Villiers R, Ittmann H & Schmitz P, 2005, Prediction of the 2004 national elections in south africa, *South African journal of science*, 101(3), pp. 157-161.
3. Greben JM, Elphinstone C & Holloway J, 2006, A model for election night forecasting applied to the 2004 south african elections, *ORiON*, 22(1), pp. 89-103.
4. Ittmann, HW, Holloway, JP & Duleni-Tlhone, N, 2017, Election Forecasting 2014 – a post-election analysis, *ORiON*, 33(1), 35-55.
5. Duleni-Tlhone, N, Holloway, JP, Ittmann, HW & Schmitz, PMU, 2018, From SA to the USA – Election Forecasting, paper accepted for publication in *ORiON*, 2018.

Other Details

Address:
Home address:
Hans W. Ittmann
Willow Green 34
Furrow Road 471
Equestria, Pretoria
South Africa, 0041

Postal address:
Hans W. Ittmann
Postent #338
Private Bag X1
The Willows, Pretoria
South Africa, 0041

e-mail address:
hittmann01@gmail.com

Contact Person:
Hans W. Ittmann