

ELECTORAL PREDICTIONS IN AFRICA:
*Predicting winners in relatively stable two-party systems,
using early and incomplete results*

Kevin S Fridy

Kevin S Fridy is an Associate Professor in the Department of
Political Science and International Studies
at the University of Tampa, Florida, USA
email: kfridy@ut.edu

ABSTRACT

In African elections, the period between polling and announcement can be protracted and tense. In the best cases, this intermission is marked by hopeful candidates urging tense supporters to stay calm. In the worst cases, such periods are used by politicians to hurl accusations of fraud back and forth to work up partisanship and devalue electoral institutions. The days between an election and its results are stressful because incomplete information about this constituency or that trickles out, but partisans have few systematic ways to compare these data with past results or exit polling, and worry that the missing data are somehow being tampered with. This paper shows how OLS regression using past results to fill in partial results can not only reduce uncertainty in the short term, but may also point out whether or not withheld results seem plausible. What began as a simple social media experiment is presented here as an elegant formula that accurately predicts outcomes across Ghana's Fourth Republic and in Nigeria's 2015 presidential election. This accuracy was achieved with as little as 10% of the results in, and extremely biased samples.

Keywords: election forecasting, Ghana, Nigeria, regression models

INTRODUCTION

After Kenyans went to the polls to vote on 27 December 2007, a tense three days would pass before the Electoral Commission declared Mwai Kibaki the winner.

This news unleashed a torrent of post-election violence that would claim more than 1000 lives and displace tens of thousands (Gibson & Long 2009). Earlier in the same year, Sierra Leone held two rounds of presidential elections, the first on 11 August and the second on 8 September. Results were not released until twelve days later in the case of the first round, and nine days later in the case of the second (Öhman 2008). A few months before that, Nigeria's Independent National Electoral Commission (INEC) took two days to release its election results, and a few more days to post slightly different results on its website (Kerr 2013). INEC Nigeria have refused to release any data on this election save for the final nationally aggregated results.

The elections described above are far from a comprehensive list. They vary in terms of location, electoral violence, outcome for incumbents, and internationally adjudicated 'free and fairness' – and paint a broad picture of what elections in Africa tend to look like. More importantly, these anecdotes illustrate a pattern that election observers on the continent, professional and amateur alike, are familiar with: polls close, days pass, election results are announced. During the 'days pass' period, results trickle in through private media outlets and more recently on social media, political operatives take to the airwaves to declare themselves winners and accuse their opposition of all kinds of foul play, and citizens grow tense.

For Ghana's 2012 elections, I wanted to see if even partial results could be used to gain some certainty for the interim between balloting and the official results. As results trickled in over the radio and internet, data from 2008 were used to fill in the gaps using OLS regression models. By the time a crowd of rowdy youth had moved from their respective party headquarters to make noise and assemble on Roman Ridge outside the Electoral Commission headquarters, I was pretty confident of what the ultimate outcome would be. Ghanaians, on the other hand, were by-and-large fed their news by the two major parties – a steady diet of biased and anecdotal evidence, insinuating their popular support and their opponents' conniving ways (Brierley & Oforu 2014; Pryce & Oidtmann 2014). Using some basic statistical tools available to most political scientists, soon after the polls closed I was able to cut through the uncertainty of post-election tension and point to a likely outcome. Divergence from this outcome would require remarkable, and potentially untoward, conditions.

As I was performing this impromptu experiment, I had no idea whether or not the models would work. Inspired by accurate predictions, I wanted to see why these simple models worked and find their limitations. Using random and non-random samples of various sizes, multiple renditions of OLS models used in the impromptu experiments, and adding four more elections, the predictive equations were put through their paces and showed themselves to be most reliable. As a final test, the model was transported from the Ghanaian context – where democracy

is relatively consolidated and the party system relatively fixed – to the Nigerian context, where past elections have been tarnished by widespread accusations of manipulation (and until 2015 there had not been a single democratic turnover of power). To bring some clarity to the confusion that thrives in the period between elections and results, my research question was: ‘Can partial results accurately predict final election outcomes?’ The answer which the various regression models provided was a tentative but hopeful ‘Yes’.

FORECASTING RESEARCH AND NEW DEMOCRACIES’ POLITICAL NEEDS

The bulk of literature on election forecasting falls into one of two categories, differentiated by the character of their independent variables. In the first category are studies attempting to predict election results using population characteristics and events as independent variables. Approval ratings, socio-economic conditions, and international security events have all been used with varying success to foretell the winners of elections before polling day (Lewis-Beck 2004; Snowberg, Wolfers & Zitzewitz 2007; Powell & Whitten 1993; Gartner, Segura & Barratt 2004). In the second category, polls of potential and/or actual voters are used to predict electoral outcomes in advance. These polls are intended to approximate electoral conditions years, months, days and sometimes hours before polls close, using randomised sampling strategies and confidence intervals to extrapolate likely results (Levy 1983; Silver 2012; Fisher et al. 2011; Walther 2015).

Unfortunately for scholars interested in forecasting elections using well-trodden paths in new democracies, polling is irregular and aggregate data are unreliable in the developing world (where most of these scholars are located). This is why novel cases outside North America and Europe are almost never incorporated into comparative studies of electoral predictors. In most African countries, pre-election polling is the realm of a few news organisations whose sampling procedures are ad hoc, and whose samples – if even described – are heavily populated by citizens living in large, easily accessible cities (Ansu-Kyeremeh 1999). Afrobarometer has conducted several rounds of public opinion surveys in dozens of countries, and some of these rounds have come close to regularly scheduled election cycles. There is a problem here, however, with a natural inflating of preferences for incumbents because of the perceived risks of being identified with the opposition (Bratton, Bhavnani & Chen 2011). The validity of aggregate data in African countries has taken much of the shine off of these reified numbers (Jerven 2013). When Nigeria became Africa’s richest economy in 2013, it did so by doubling its GDP without changing the living standard of a single Nigerian (Magnowski 2014). If rebasing calculations can change such a fundamental piece of economic data in

a profound way, it would take a major leap of faith to rely on this – or even less scrutinised – aggregate data to build a predictive model in which one could have any confidence.

Although recording of aggregate characteristics and pre-election polling remains poor and uneven in the developing world because of cost and infrastructural realities, fairly substantial electoral data do exist. These data inform the aforementioned models based on aggregate characteristics and events, as well as those based on polling. They provide the historical equilibrium at which contemporary occurrences and trends can tug and pull toward one party or another. In a context where elections occur days or sometimes even weeks before winners are announced officially by electoral commissions, such data can usefully be put to forward-looking forensic purposes (Geben, Elphinstone & Holloway 2006¹).

In the African context, elections usually follow a similar script. Shortly after polls close, political operatives take to the airwaves urging partisans that they have already won the election and warning that vigilance is necessary because their opponents are trying to steal victory under cover of night. In some cases this behaviour results in little more than chest-pounding and minor bouts of youth disorderliness, but in other cases the tension of this period of uncertainty boils over into mass violence. Partial election results can show how likely particular electoral outcomes are, and if there is an unexpected outcome once the results are declared, results that deserve closer inspection can be more easily identified. This would help in search of the causes, whether legitimate or illegitimate, of these discrepancies.

GHANA 2012: AN IMPROMPTU EXPERIMENT

When Ghanaians went to the polls in 2012, I watched the campaign from afar. Reading newspaper reports on the internet became a daily ritual, and unlike in past elections, I added my say to the news on social media. President Mills had died just four and a half months earlier. His death meant the election would pit an incumbent, John Mahama, who had never run a national campaign on the top of the ticket, against a challenger, Nana Akufo-Addo, who had won the plurality of votes in 2008's first round – and barely lost in its second, and who had already been campaigning for several months. Because of these unusual circumstances,

1 South Africa's Council for Scientific and Industrial Research (CSIR), to which Geben, Elphinstone and Holloway are attached, has had similar ideas. Their analysis is the only regular systematic forecasting done on the continent. It has a very good track record, but requires a far more sophisticated model than that presented here, and is much more data intensive. Both these characteristics present substantial problems in most other African countries where data and econometric skills are more limited than in South Africa.

partisans on both sides were confident about their chance of victory. With roughly one in every 15 Ghanaians having a Facebook page, for people like me who had many Ghanaian friends, the campaign took centre stage in news feeds (Fripp 2014).

Polls for the first round of voting were open on 7 December 2012, but results trickled in slowly. Ghana's Electoral Commission (EC) does not issue certified results until they have all been compiled, leaving news outlets to gather bits and pieces of information from their contacts within the EC and reporters at individual polling stations, where ballots are physically counted in view of the public. Because the EC is reluctant to publish results at the polling station level ($N > 25\,000$), for comparison's sake the results are worthless, except in the eyes of residents who attended the counting in previous elections and remember the results. The news outlets rely on constituency-level election observers situated at compiling centres to push out the results to the general public before they are formally certified and released by the EC days later.

By 13h00² on 8 December 2012, only 22 of the country's 275 constituencies had been reported by Joy FM.³ With slightly less than 10% of constituencies tallied, I posted the following comment approximately 18 hours after polls closed:

I just used Joy FM's incomplete data to predict the eventual winner. I ran two regressions, one using 2008 New Patriotic Party (NPP) results to predict 2012 NPP results based on the incomplete data and the other doing the same for the National Democratic Congress (NDC). I got the following two linear models as a result: $NDC_{12} = .067 + .926(NDC_{08})$ and $NPP_{12} = .011 + .943(NPP_{08})$. This means that a constituency where the NPP and NDC won 50 percent of the vote in 2008, they are polling about 53 percent for the NDC and 48 percent for the NPP. Very preliminary figures look good for John Mahama.⁴

Substituting actual values from 2008's first round of presidential votes instead of the hypothetical 50% mentioned in my post, these models predict a victory for the NDC candidate, Mahama, with 51.07% of the vote. Akufo-Addo is predicted to poll 47.43% for the NPP. Actual numbers after all the votes were tallied were 50.70% for Mahama and 47.74% for Akufo-Addo. Hence, this early rudimentary model was accurate to within half-a-percentage point on both accounts.

2 All times are reported on the 24-hour clock in Greenwich Mean Time (GMT) hours.

3 Joy FM (2012) is an independent radio station based in Accra. Founded in 1995, it is Ghana's first privately licensed radio station and part of the Multimedia Group media conglomerate. Though I checked other sources, Joy FM tended to be the quickest to release results and their results matched those of other media houses (which followed Joy FM's) lead.

4 I present these Facebook posts unedited for content, but have corrected typos and grammatical errors and spelt out abbreviations.

Two-and-a-half hours later, approximately 15h30 the day after the election, the results for nearly a quarter of the constituencies had been reported. I reran my analysis with the 63 known constituency results. These updated models predicted the final outcome of the race to be Mahama with 50.50% and Akufo-Addo with 47.63%. This proved to be another successful experiment, as both totals were slightly closer to the actual numbers than in the previous model. The first model used only 10% of the results, whereas the second model used the newly available 23% of results.

Between this second post and the following post, six hours passed. I realised that my sample was likely not random. Some regions had compiled and reported nearly all their results but others were missing nearly all their data. Given what I know about the varying preferences of Ghana's regions (Fridy 2007), I worried that a disproportionate number of one party's strongholds pending could make my predictions highly inaccurate. Dummy variables were created to address this potential deficiency.

Using the same constituencies I last reported on I wanted to see if this election is shaping up to be more or less polarized than the last election so I threw in a dummy variable for the NPP strongholds (Ashanti and Eastern region) and NDC strongholds (Volta and three Northern regions) and got the following lines: $NPP12 = .027 + .876(NPP08) + .045(NPPRegionDummy) - .011(NDCRegionDummy)$ and $NDC12 = .120 + .851(NPP08) - .052(NPPRegionDummy) + .012(NDCRegionDummy)$. This is interesting stuff! The model suggests that NPP and NDC did similarly well in the Ashanti and Eastern regions as they did in 2008. In the mythical Ashanti/Eastern constituency where Mills and Akuffo-Addo got 50 percent in 2008 the model predicts the NDC will pick up just shy of 50 percent in 2012 and NPP just over 50 percent in 2012. For Swing and NDC regions, however, there has been a lot of movement. In these areas the prediction is for Mahama to pick up about 55 percent of the vote and Akuffo-Addo about 46 percent.

Unlike previous models, these regressions were designed to predict individual constituencies' outcomes and general trends, and do not translate into a national prediction without additional inputs.

Just after midnight on Sunday my most complex model was posted. This model had nearly half the total constituency results available as actual data. The model, which used past performance and dummy variables controlling for strongholds, was used to predict the missing data. Predicted percentages were

then multiplied by the total number of actual votes from 2008 to approximate total votes from each constituency for the two major parties in 2012.

Last prediction of 2012. With information on just under half of the constituencies (N=130), I get the following regression equations. I must say these have been pretty consistent since we had only a couple of dozen constituencies which impressed me at least. $NPP12 = .052 + .816(NPP08) + .057(NPPRegionDummy) - .042(NDCRegionDummy)$ and $NDC12 = .166 + .773(NDC08) - .071(NPPRegionDummy) + .045(NDCRegionDummy)$. I ran every unknown constituency through these formulas and then added the knowns in and came up with the following: John Mahama (51.6 percent) Nana Akufo-Addo (45.8 percent). I think the region dummies, even with their relatively low coefficients, are necessary because most of the unknown votes are in relatively heavy NDC areas. I am pretty comfortable at this point predicting a John Mahama one touch.⁵ We shall see tomorrow or Monday how close these numbers are to the tallies published by the EC.

This final model adjusts for varying constituency size, which is a major improvement on the previous models. For example, Sekyere Afram Plains in Ashanti Region has just over 12 000 registered voters, whereas Ketu South in Volta Region and Ledzokuku in Greater Accra have more than 127 000 voters each. The ability to account for varying size greatly enhanced the face validity of the final model. As far as predictive values go, however, the predicted results were close to the actual results, but not as close as the results of the simpler (earlier) models – which had access to less actual data.

SYSTEMATICALLY EXPLORING RESULTS OF IMPROMPTU EXPERIMENT

Buoyed by the promising results from these impromptu experiments, I set out to test the models described above more systematically, in hopes of finding out whether I had gotten lucky or was merely onto an efficient method of decreasing anxiety between elections and results. If the model works across elections, I wanted to figure out how early in the process of reporting results it can begin to produce reliable projections. I also wanted to figure out which parts of the above regression models were necessary and which parts were superfluous.

⁵ 'One touch' is the phrase Ghanaians use to describe a candidate obtaining a majority of the votes in the first round of presidential balloting, thus avoiding a potential second round contest.

My final experimental model had included not only previous results, but also dummy variables for traditional strongholds and controls for constituency size. For parsimony, I needed to find out which of these variables could be trimmed from future models. First-round presidential contests were used as test cases, because that yielded six comparable cases in the Fourth Republic. In 1992 Ghana had 200 constituencies, in 2004 it had 230 constituencies, and in 2012 the constituencies were increased to 275. Fortunately for the project at hand, newly minted constituencies are always the result of a single old constituency being segmented into two or three new constituencies, making it a simple addition problem to aggregate up to 200 like units.

The test began with ten random selections of 10% (N=20) and 25% (N=50) of constituencies. I wanted to explore the benefits of increasing the sample size to see how early in the process of released unofficial results the forecasting models begin to work. Though I anticipated that Ghana's strong two-party system would be largely unaffected by the difference, I was interested in testing models using one and two parties' previous results. In a situation where there is more party volatility and/or more significant parties, it could be useful to include all the parties gaining a substantial portion of the vote in the model. Finally I wanted to see how important it is to weight constituencies, given their vastly different numbers of registered voters. I began to venture into this territory in later versions of my impromptu experiments, and given the known size of each constituency's voters' roles, it was easy to take population sizes into account.

Table 1 displays the results of these tests. The first striking feature of these regressions that use complete 2008 election results and partial 2012 results to predict the outcome of the 2012 election are the coefficients of determination. Not surprisingly, the adjusted R² values were slightly higher when I used a quarter of the 2012 results than when I used a tenth. For both cases, however, the mean R² was quite high, with the lowest adjusted R² being .852. This finding suggests – as one would expect given the recurring patterns in Ghanaian elections – that past election results are really good predictors of future election results.

Adding data from both the NDC and NPP parties⁶ does not produce consistently more accurate models than using data from only the party one is trying to predict.⁷ In the eight models shown in Table 1 (one run for the NPP and one for the NDC), there are 16 generic regression equations. Two-party models perform slightly better in half the cases, one-party models in nearly 40% of the cases, and there is a tie in one case. Even when there is a slight difference, however, it never surpasses 0.003%. Weighting regressions by the size of constituencies does

6 Party A 2012 = b + m(Party A 2008) + m(Party B 2008)

7 Party A 2012 = b + m(Party A 2008)

Table 1
Random tests using 2008 first-round presidential results to predict 2012 results

(a) 10 random tests from 10% of constituencies				
Independent variables		Predicted average	Predicted maximum	Predicted minimum
1 party	NDC (Mahama)	52.2%	54.3%	50.4%
	NPP (Akufo-Addo)	46.5%	48.3%	44.6%
2 parties	NDC (Mahama)	51.9%	53.7%	50.1%
	NPP (Akufo-Addo)	46.4%	48.4%	44.7%
1 party weighted	NDC (Mahama)	51.9%	53.5%	50.3%
	NPP (Akufo-Addo)	46.8%	48.4%	45.4%
2 parties weighted	NDC (Mahama)	51.7%	52.8%	50.3%
	NPP (Akufo-Addo)	46.8%	48.2%	45.4%
(b) 10 random tests from 25% of constituencies				
Independent variables		Predicted average	Predicted maximum	Predicted minimum
1 party	NDC (Mahama)	51.7%	52.2%	50.8%
	NPP (Akufo-Addo)	46.7%	47.4%	46.1%
2 parties	NDC (Mahama)	51.6%	52.1%	50.8%
	NPP (Akufo-Addo)	46.8%	47.4%	46.2%
1 party weighted	NDC (Mahama)	51.3%	52.1%	50.4%
	NPP (Akufo-Addo)	47.2%	47.9%	46.7%
2 parties weighted	NDC (Mahama)	51.4%	52.0%	50.6%
	NPP (Akufo-Addo)	47.1%	47.7%	46.6%

Source: Ghana Electoral Commission (n.d.)

Electoral Commission Certified Results: Mahama 50.7%, Akufo-Addo 47.7%

Notes: For the 10% tests, adjusted R² varied from .973 to .852. The average adjusted R² for the 10% tests was .939; for the 25% tests, adjusted R² varied from .982 to .884. The average adjusted R² for the 25% tests was .956.

give consistently better results. In all instances weighted regressions were more accurate than their non-weighted counterparts, although this enhanced accuracy advantage was slight and never rose above 0.004%. Random tests conducted knowing a quarter of the 2012 results were, as anticipated, better than similar tests conducted knowing only a tenth of the results. For both the NDC and NPP, more results led to a prediction that was roughly 0.004% better on average across the four models and ten samples.

More important than the differences between the models were the similarities. Table 1 depicts the results of four models with a total 20 of random samples. In total, 80 tests were performed. Not a single test predicted anything but the accurate outcome. In no test was Mahama predicted to win less than 50.1% of the vote. In no test was Akufo-Addo predicted to win more than 48.4% of the vote. Some model tests were slightly better and some slightly worse than the impromptu experiments described above, but on the whole they were much on par.

The random samples drawn to perform the preceding tests are likely not representative of real world returns. Even under conditions where everything is above board and nothing untoward is happening, rural areas can take longer to report because of logistical difficulties, and large constituencies can be delayed because of long lines at poll closings. Weather can cause delays in some regions but not others, and if these regions differ significantly in voting patterns compared with the rest of the country, premature predictions will paint an inaccurate picture.

To see how the best model in Table 1 (two-parties weighted) works under situations that are not random, the 200 Ghanaian constituency units are divided into three subsets: NDC strongholds (N=56), NPP strongholds (N=62), and competitive (N=82). Strongholds are defined by a single party winning by 10 percentage points or more in the last three first-round presidential elections (2000, 2004 and 2008).

Table 2 shows the results for prediction regressions using 2008 first-round presidential results and partial results from the 2012 first-round presidential contest to predict the election's final outcome. For NDC and NPP strongholds, as well as for competitive districts, the regression models accurately predict a first-round Mahama victory. The adjusted R^2 values in these models are less robust than those in the random models, which suggests that the samples are problematic, but are still quite impressive as they never fall below .750. This finding suggests that elections, at least those in Ghana, work a bit like a dimmer switch. Across a diverse array of constituencies, the switch gets turned a little in favour of one party or another. Perhaps it is economics, social policy, pro- or anti-incumbent sentiments, or some less well-studied *zeitgeist* that does the turning, but whatever the independent variable, it affects the results across constituencies similarly. Put another way, in 2012 the NDC did a few percentage points better in NPP

strongholds than in 2008. This statement need not be adjusted to describe NDC strongholds and competitive constituencies. As a result, this finding suggests the laborious models of my later impromptu regressions, which segmented Ghana into partisan blocs through the use of dummy variables, were unnecessary. The simpler models in Table 1 work well even for the most extreme non-random samples.

Having shown that a relatively simple regression model works well in filling out missing 2012 election results, I was interested to see if 'election 2012' was an anomaly. Using the same 200 constituency units sampled and the same randomly drawn samples of a quarter of constituency units in the models above, first-round presidential elections in 1992, 1996, 2000 and 2004 were used to fill in the gaps from missing election results for 1996, 2000, 2004 and 2008 respectively. If the prediction regressions work in these tests, every single Fourth Republican election will demonstrate the utility of this model. Table 3 displays these results.

For every election under consideration (N=4) and both of the major political parties (N=2), these random samples (N=10) yield accurate results when run through the regression model. That is 80 regressions without a single inaccurate result. Rawlings wins handily for the NDC in 1996; Kufuor needs a second round to beat Mills in 2000 for the NPP, but clears the 50% hurdle in 2004's first round; and Mills gives the presidency back to the NDC in 2008, but needs a second round to defeat his opponent by a very narrow margin. Two of these elections preserve the status quo and two push Ghana past Huntington's (1991) vaunted 'two turnover test.' Two elections are won by the NPP and two by the NDC. Two elections are won with a single round and two require a run-off. Despite these different electoral

Table 2
Test using 2008 first-round presidential results to predict 2012 results in strongholds

	Units	NDC (Mahama)	NPP (Akufo-Addo)	Adj. R ²
NDC Stronghold	56/200	53.3%	44.3%	.864 (NDC) .813 (NPP)
NPP Stronghold	62/200	50.7%	48.0%	.790 (NDC) .786 (NPP)
Competitive	82/200	51.8%	46.6%	.765 (NDC) .777 (NPP)

Source: Ghana Electoral Commission (n.d.)

Electoral Commission Certified Results: Mahama 50.7%, Akufo-Addo 47.7%

Notes: Strongholds are defined as constituency units (N=200) where a party won by at least 10% of the vote in 2000, 2004 and 2008 first-round presidential elections.

contexts, the use of past election results and partial contemporary election results to fill in the gaps of missing election data and predict final results provides a consistently accurate approach.

Table 3
Random tests using Ghana's previous election first-round presidential results to predict future results

		Actual	Predicted		
			Average	Maximum	Minimum
1996	NDC (Rawlings)	57.4%	57.0%	58.3%	54.1%
	NPP (Kufuor)	39.7%	39.6%	40.7%	36.5%
2000	NDC (Mills)	44.5%	45.7%	47.1%	43.9%
	NPP (Kufuor)	48.2%	46.8%	47.6%	46.0%
2004	NDC (Mills)	44.6%	45.0%	46.8%	43.1%
	NPP (Kufuor)	52.5%	52.3%	54.2%	50.4%
2008	NDC (Mills)	47.9%	48.8%	49.9%	47.1%
	NPP (Akufo-Addo)	49.1%	48.0%	49.0%	46.9%

Source: Ghana Electoral Commission (n.d.)

Notes: All tests had 10 renditions with different randomly selected 25% samples. The regression equation was $Party A_t = b + Party A_{t-1}(m_a) + Party B_{t-1}(m_b)$ and was weighted by the number of registered voters. Adjusted R^2 across all election predictions varied from .978 to .761. The average adjusted R^2 overall was .904. The average adjusted R^2 for the NDC was .894 for 1996, .937 for 2000, .870 for 2004 and .922 for 2008. For the NPP the R^2 was .871 for 1996, .942 for 2000, .895 for 2004 and .901 for 2008.

NIGERIA 2015: A NEW TEST FOR THE PREDICTION MODEL

When Nigerians went to the polls in 2015, I used the election as a test of these simple regression prediction techniques. There was some consistency between the 2011 and 2015 elections, with Goodluck Jonathan and Muhammadu Buhari being the presumed main contenders in both. Although it is possible for this technique to work in cases where candidates and parties change markedly from election to election, it would require a certain expectation of stability in social cleavage and political party interaction. While there was consistency in the top two contenders, Nigeria's 2015 race differed from the 2011 contest in ways that make it markedly different from the Ghanaian case. Though Nigeria's Fourth Republic has been holding democratic contests since 1999, only seven years less than Ghana's Fourth Republic, the case for democratic consolidation is much

harder to make in Nigeria. Prior to 2015 Nigeria had not seen a single democratic turnover of power, opposition parties rise just in time for an election only to disappear before the next, and reported election results are suspect (Omotola 2010).⁸ In other words, Nigeria is a novel enough case for the prediction model to demonstrate its more general reliability.

Nigeria's Independent National Electoral Commission (INEC) was not reporting partial results in the 2015 contest. INEC's policy is not to publish results until the counting is finished and results certified. Electoral Commissioner Jega went so far as to warn media outlets that published partial results that 'Only INEC is empowered by law to announce results and it is an offence for anyone to preempt the Commission in this regard' (The Citizen 2015). One of the accused scofflaws was *Sahara Reporters* (SR), a New York-based online news agency that has been described as 'Africa's WikiLeaks' (Shenon 2010). Outside of Jega's reach, SR published several updates on election results prior to INEC's official announcement (Sahara Reporters 2015). Muhammadu Buhari, in an election that took place on Saturday 28 March, would not be declared the winner by INEC until the early morning hours of Wednesday 1 April. SR published partial results of all states (N=37)⁹ just before 01h00 on Monday 30 March (SR Result 1), again just after 09h30 in the morning (SR Result 2), and again just after 22h00 that evening (SR Result 3). This data gives the aforementioned forecasting model the data it needs to be tested.

The SR results, advertised as provided by a 'credible source,' are uneven and at least occasionally erroneous. Jonathan's strongholds in the southeast were quite late in reporting to INEC and being reported by SR. In SR Result 1, Buhari had already collected more than a quarter of his total votes compared to Jonathan's seventh. By SR Result 3 this gap had narrowed, but Buhari had 92% of his final votes counted compared with Jonathan's 83%. In three instances for Buhari and three for Jonathan, votes are slightly over-reported by SR. In none of these six instances does a candidate receive more than 108% of their final vote. The rarity of occurrences, small margins, and fact that Buhari and Jonathan have similarly over-reported votes indicate random error in the model, rather than systematic error.

Less than an hour after SR posted its first preliminary results, I ran the numbers through the regression model. I posted the following comment on Facebook on Monday 29 March at 01h22. At that point polls had been closed for over a day and SR had partial results for only 30 of Nigeria's 37 states. Only Ekiti had complete results, and Buhari had collected 27% of his votes while Jonathan had collected 14%.

8 In Yar'Adua's 2007 election, results were so problematic INEC has refused to this day to publish results at the subnational level.

9 For the purposes of this paper, Abuja (Nigeria's Federal Capital Territory) is treated as a state unit.

I put this data into an OLS regression using Goodluck Jonathan and Muhammadu Buhari's results from 2011 to fill in the blanks in the 2015 data. Assuming this *Sahara Reporters'* data is legit and the states reporting are not significantly different from those that are not reporting, the model says Buhari will get about 60 percent of the vote and Jonathan about 40 percent. If you are interested in replicating I used the formula $\text{Candidates' 2015 vote} = b + m(\text{Candidates 2011 vote})$ and weighted for number of 2011 voters.

With $N=30$ and less than a quarter of votes reported (from what could be a random or non-random sample of polling stations), this prediction was ten points closer to the actual results than the raw percentages being reported by *SR*. The latter source said Buhari would claim just shy of 70% of the vote and Jonathan just over 30%.

Table 4
Prediction models versus reported results in Nigeria's 2015 presidential election

% of votes reported	Source of Results	APC (Buhari)	PDP (Jonathan)
21%	<i>Sahara Reporters' raw</i>	69.4%	30.6%
	Model prediction	58.3%	41.7%
40%	<i>Sahara Reporters' raw</i>	72.8%	27.2%
	Model prediction	56.3%	43.7%
88%	<i>Sahara Reporters' raw</i>	57.2%	42.8%
	Model prediction	53.6%	46.4%
100%	INEC final results	54.5%	45.5%

Source: *Sahara Reporters (2015) and INEC (2015)*

Notes: *Sahara Reporters* reported only results for the top two contenders. Because candidates other than Buhari and Jonathan received only 1.08% of the vote, these numbers do not differ vastly from results that include all the candidates (but they are slightly different). To regularise the units, INEC's final results are calculated without the minor candidates as well. Model prediction uses the partial results from *Sahara Reporters* as the dependent variable and INEC results (2011) as the independent variables. The regression equations were $\text{Buhari 2015} = b + \text{Buhari 2011}(m) + \text{Jonathan 2011}(m)$ and $\text{Jonathan 2015} = b + \text{Buhari 2011}(m) + \text{Jonathan 2011}(m)$. The regressions were weighted by the number of votes in a state in 2011. $N=37$, and adjusted R^2 across all election predictions ranged from .788 to .907. Average adjusted R^2 overall was .829.

Table 4 shows how the regression forecasting model fared compared with *SR*'s raw percentages. As was the case when I worked with data from just 21% of the votes reported, once 40% and then 88% of the votes were reported, the regression forecasting model predicted results that were increasingly closer to the actual final results, compared with the raw percentages offered by *SR*. Whereas raw percentages reported by *SR* put Buhari ahead of Jonathan by 40 percentage points, the regression forecasting model predicted a much tighter race, with Buhari leading by only 15 percentage points. As more results trickled in, *SR*'s percentages came closer to the actual final INEC results but never got as close as the regression forecasting model's results. On average, raw percentages reported by *SR* were off by 12%, suggesting Buhari was due for a much bigger victory than he actually achieved. The regression forecasting model using Buhari and Jonathan's results from 2011 and weighting for state size was only 2 percentage points off, on average. A Buhari victory was still predicted, but one that was more closely tied to actual results.

CONCLUSION

In the case of Ghana's Fourth Republic, past election results are remarkably good predictors of their successors, assuming very little data are collected. Only knowing 10% of the 2012 results, those past election results reliably predicted a Mahama one-touch victory. While the Ghanaian media presented anxious voters with a story that resembled 'it's anybody's race', pastors, mallams, musicians, actors and footballers sought to outdo each other by praying for peace, and signalled to citizens that an uncertain future awaited that might disappoint – or worse, might be cause for alarm. Yet less than a dozen hours after polls closed the race was not 'anybody's to win.' At that point all signs pointed to Mahama's eventual victory, and anything contrary to that result would either indicate a rare anomaly or fraud.

Lessons learned from the Ghanaian case proved useful in interpreting partial results from Nigeria's 2015 election. A day after polling ended and with only 21% of the vote known, there were strong indicators that Buhari would win the historic elections in one round. This prediction was surprisingly accurate despite the fact that Jonathan's home region was late in reporting. Nigerians would be led to believe the election was up for grabs for another two-and-a-half days. Elections breed anxiety and there are ways to mitigate the length of this period of marked tension.

Despite these two successful applications, some substantial questions remain unanswered. First, how scalable are the tests offered here? Ghana has a stable two-party system. Although Nigeria's party system is decidedly less stable, the

top two presidential candidates in 2011 and 2015 were identical. In countries where parties change names with the seasons, and candidates regularly come and go, this system of election predictions might – or might not – work. A dearth of subnational election results on the continent makes testing these hypotheses difficult, but a more diverse set of trials is called for. Only with more real-world tests can we find the model's limits.

Second, what is the utility of the knowledge gained from these predictions? In most cases we are talking about predictions that preceded the certified results by only a few days. If these days are the only pay-off, the regression models amount to not much more than cheap parlour tricks. There is no systematic evidence suggesting that the 'wait and see' post-election period leads to more electoral violence or less respect for electoral institutions compared with the more instant gratification of quicker results. There is, however, plenty of anecdotal evidence to suggest that tensions do run high during this period of purgatory, and competing narratives begin to harden.

— REFERENCE —

- Ansu-Kyeremeh, K 1999, 'The challenges of surveying public opinion in an emerging democracy', *International Journal of Public Opinion Research*, vol. 11, no. 1, pp. 59–74.
- Bratton, M, Bhavnani, R & Chen, T-H 2011, 'Voting intentions in Africa: ethnic, economic or partisan?', *Afrobarometer Working Paper*, vol. 127.
- Brierley, S & Ofori, G 2014, 'The presidential and parliamentary elections in Ghana, December 2012', *Electoral Studies*, vol. 35, pp. 362–405.
- Electoral Commission of Ghana 'Election Results', *ECG*, accessed from <http://www.ec.gov.gh/election-results/>
- Fisher, SD, Ford, R, Jennings, W, Pickup, M & Wlezien, C 2011, 'From polls to votes to seats: forecasting the 2010 British general election', *Electoral Studies*, vol. 30, no. 2, pp. 250–257.
- Fridy, KS 2007, 'The elephant, umbrella, and quarelling cocks: disaggregating partisanship in Ghana's Fourth Republic', *African Affairs*, vol. 106, no. 423, pp. 281–305.
- Fripp, C 2014, 'Top 10 African countries on Facebook', *IT News Africa*, accessed from <http://www.itnewsafrika.com/2014/02/top-10-african-countries-on-facebook-2/>
- Gartner, SS, Segura, GM & Barratt, BA 2004, 'War casualties, policy positions, and the fate of legislators', *Political Research Quarterly*, vol. 57, no. 3, pp. 467–477.

- Gibson, CC & Long, JD 2009, 'The presidential and parliamentary elections in Kenya, December 2007', *Electoral Studies*, vol. 28, no. 3, pp. 497–502.
- Greben, JM, Elphinstone, C & Holloway, J 2006, 'A model for election night forecasting applied to the 2004 South African elections', *ORiON: The Journal of ORSSA*, vol. 22, no. 1, pp. 89–103.
- Huntington, SP 1991, *The Third Wave: democratization in the late twentieth century*, University of Oklahoma Press, Norman, OK.
- Independent National Electoral Commission, Nigeria 2015, 'Results for 2015 presidential general elections', *INEC*, accessed from http://www.inecnigeria.org/?page_id=31
- Jerven, M 2013, *Poor numbers: how we are misled by African development statistics and what to do about it*, Cornell, Ithaca, NY.
- Joy FM 2012, 'Ghana elections 2012', *My Joy Online*, accessed from <http://elections.myjoyonline.com/>
- Kerr, N 2013, 'Popular evaluations of election quality in Africa: evidence from Nigeria', *Electoral Studies*, vol. 32, no. 4, pp. 819–837.
- Levy, MR 1983, 'The methodology and performance of election day polls', *Public Opinion Quarterly*, vol. 47, no. 1, pp. 54–67.
- Lewis-Beck, MS 2004, 'Election forecasting: principles and practice', *British Journal of Politics and International Relations*, vol. 7, pp. 145–164.
- Magnowski, D 2014, 'Nigerian economy overtakes Africa's on rebased GDP', *Bloomberg*, accessed from <http://www.bloomberg.com/news/2014-04-06/nigerian-economy-overtakes-south-africa-s-on-rebased-gdp.html>
- Öhman, M 2008, 'The 2007 presidential and parliamentary elections in Sierra Leone', *Electoral Studies*, vol. 27, no. 4, pp. 764–768.
- Omotola, JS 2010, 'Elections and democratic transition in Nigeria under the Fourth Republic', *African Affairs*, vol. 109, no. 437, pp. 535–553.
- Powell, GB & Whitten, GD 1993, 'A cross-national analysis of economic voting: taking account of the political context', *American Journal of Political Science*, vol. 37, no. 2, pp. 391–414.
- Pryce, P & Oidtmann, R 2014, 'The 2012 general election in Ghana', *Electoral Studies*, vol. 34, pp. 330–334.
- Sahara Reporters 2015, '#NigeriaDecides', *Nigerian Community Online News*, accessed from <http://saharareporters.com/nigeriadeclides>
- Shenon, P 2010, 'Africa's WikiLeaks', *The Daily Beast*, accessed from <http://www.thedailybeast.com/articles/2010/08/12/sahara-reporters-uncovering-nigerias-corruption.html>
- Silver, N 2012, 'FiveThirtyEight's 2012 forecast', *New York Times*, accessed from <http://fivethirtyeight.blogs.nytimes.com/fivethirtyeights-2012-forecast/>
- Snowberg, E, Wolfers, J & Zitzewitz, E 2007, 'Partisan impacts on the economy:

evidence from prediction markets and close elections', *Quarterly Journal of Economics*, vol. 122, no. 2, pp. 807–829.

The Citizen 2015, 'Presidential poll: Jega scores card readers performance high', *The Nigerian Voice*, accessed from <http://www.thenigerianvoice.com/news/174498/1/presidential-poll-jega-scores-card-readers-perform.html>

Walther, D 2015, 'Picking the winner(s): forecasting elections in multiparty systems', *Electoral Studies*, vol. 40, pp. 1–13.