Forecasting Performance of Arma and Arfima Models in Short Time Series: An Analysis of Kenya Political Opinion Poll Data

Otieno M. O.
Department of Mathematics
Egerton University
P.O Box 536, Egerton, Kenya.

Mwangi J. W.
Department of Mathematics
Egerton University
P.O Box 536, Egerton, Kenya.

Islam A.S.
Department of Mathematics
Egerton University
P.O Box 536, Egerton, Kenya.

Abstract

The popular approach to modeling time series data is to apply the Box-Jenkins approach of ARMA or ARMA depending on whether the series is stationary or non-stationary. This approach is based on the assumption that the sample is large. If such series display long memory property, then the forecast values based on ARMA model may not be reliable. In case of short time series data, one cannot rely on estimation techniques based on the asymptotic theory. This calls for use of appropriate estimation techniques in order to come up with models that can capture the short time series properties and thus be adequately used for prediction and forecasting without loosing the principle of parsimony. This study focuses on fitting appropriate ARMA and ARFIMA models for short time series and measuring the forecast performance of the fitted models. The percentage political popularity ratings series for three presidential candidates in Kenya's General Elections for the year 2007 are used. A model-selection strategy based on the corrected Akaike Information Criterion (AICc) is adopted to determine the correct model specification. Exact maximum likelihood estimation method is used to estimate the model parameters. RMSE is used to evaluate the forecast performance of the model. ARFIMA models are found to represent and forecast the short time series polls data better than the ARMA models.

Keywords: Opinion polls, Short time series analysis, exact MLE, RMSE, ARFIMA, ARMA.

1.0 Introduction

The time series analysis of political support for Kenya's three main presidential contenders in the 2007 general elections is of some interest. This is because Kenya only regained democratic multiparty system in 1991 after many years of single party system. For this reason, the length of the available opinion polls data is much shorter than in other countries with a much longer and mature democratic tradition. This study seeks to find out whether the statistical properties of the opinion polls in Kenya conform to patterns found for those other countries, despite its younger democracy. The study further seeks to find out the most adequate models for these data and the forecasting performance of the fitted models.

Over the last few years there has been growing interest among political scientists in applying time series techniques to analyze the statistical properties of aggregate political popularity data in various formats such as approval levels and partisanship measures.

For example, [1,2], [3,4], [5], [6], [7], [8] and various articles included in a special issue of Electoral Studies in the year 2000 report evidence for the United States, United Kingdom, Spain, Sweden, Finland and several other OECD countries. These studies indicate that the time series of poll ratings in those countries are well modeled by fractionally-integrated processes which present high persistence but that eventually revert to their mean.

A fractionally integrated process is one that exhibits long memory, with persistent local trends, but which nonetheless eventually reverts to the mean, [3]. The degree of persistence is measured by a real-valued parameter d, lying on the unit interval. At one extreme end, $d = 0$ represents the short memory case. If $d > 0.5$, the process is not wide-sense stationary, having infinite variance. And at the other extreme end, $d = 1$ corresponds to the ordinary integrated process, familiarly known as a random walk, which is well known not to revert to the mean but to eventually wander arbitrarily far from the starting point. In [3] each derives variants of a model that explains the fractional property as consequence of aggregating heterogeneous poll responses. In these models, d measures the distribution in the voting population of a certain individual characteristic, which [2] call persistence of party identification, and [3] describe in terms of commitment versus pragmatism. These are attributes that might well be supposed to depend on the political culture, traditions, and attitudes of voters, and to vary from one party to another, and also from one country to another [3].

The rest of this paper is structured as follows. Section 2 summarizes the micro foundations of the model of voting intentions proposed by [3], using the model for partisanship proposed by [9. We emphasize the key assumptions which would give rise to an ARFIMA process as an appropriate model governing the time-series behavior of aggregate poll series. Section 3 explains both the estimation and testing approaches used in this paper. We report the results and the conclusion in sections 4 and 5 respectively.

2.0 Microfoundations of the popularity model

The [3] model is based on the idea that voters fall into two categories; the 'committed' and the 'floating' voters. Support of the 'committed' voters is determined mainly by conviction or group solidarity, and so is relatively insensitive to the current performance of the party. The 'floating' voters are more pragmatic, and their support is driven mainly by performance. It follows that the future voting behavior of the second group is typically less predictable from current behavior than the first group. The degree of persistence of aggregate support depends on the distribution of these attributes in the voter population.

The [3] model assumes that the log-odds in favor of voter i supporting a given party is described, apart from a deterministic component, by an autoregressive process driven by news. In other words, if p_t^i represents the probability of voter i supporting the party at time t then

$$\log \frac{p_t^i}{1-p_t^i} = C^i + y_t^i \qquad (1)$$

where

$$y_t^i = \alpha^i y_{t-1}^i + \varepsilon_t^i \qquad (2)$$

The term C^i is time-varying and it captures the effect of the election cycle. Equation (2) measures the degree of persistence of party support in the face of 'news', whose effect on the individual is measured by ε_t^i

Assuming ε_t^i to be a serially uncorrelated process, the case $\alpha^i = 1$ in (2) corresponds to a random walk process, which evolves with high probability towards $+\infty$ and $-\infty$, so that the probability of support p_t^i tends to unity or zero under (1). Thereafter, it changes only rarely. This represents the behavior of committed voters. On the other hand $\alpha^i < 1$ implies a reversion to mean, and hence of p_t^i migrating (in the particular case $C^i = 0$) to 1/2, in the absence of news. Because of the nonlinearity of the logistic transformation, support is also a lot more volatile in this case, in the face of the same news, than it is in the unit root case. This case represents the shorter 'memory' of pragmatic voters. The α^i are assumed to be distributed in the voting population over the interval [0, 1] according to the beta(u, v) density, where u and v are constant parameters, and $0 < v < 1$.

For a suitable choice of v, this distribution can concentrate a significant part of the probability mass very close to 1. Since the beta is a very flexible functional form, the distribution can assume a range of shapes on the rest of the interval, depending on the parameters. It can be approximately uniform.

Let $\overline{X}_t$ represent the arithmetic average of N independent binary (0-1) opinion poll responses, sampled from the population at time t, such that $100\overline{X}_t$ is the usual percentage support measure. Consider the time series properties of log[$\overline{X}_t/(1-\overline{X}_t)$] when t represents a succession of time periods (monthly or quarterly). [3] show that this variable converges in probability as $N \to \infty$ to the same limit as $\overline{C}+\overline{y}_t$, the mean of the right hand side of (1), where $\overline{C}$ is converging to a constant and

$$\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_t^i$$

$\bar{y}_t$ is a random variable in the limit, being a function of news variable that all voters observe, although the individual effects are averaged out. The key result, due to [10], is that under a beta(u, v) distribution for the α^i, the time series representation of $\bar{y}_t$ approximates (large N) to a process of the form

$$\bar{y}_t = \sum_{k=0}^{\infty} \alpha_k \bar{\varepsilon}_{t-k}$$

where $\alpha_k = O(k^{-v})$, and $\bar{\varepsilon}_t$ is a shock process depending on news. This says that averaging a mixture of stable autoregressions and near-unit root processes yields in the limit a moving average process whose coefficients decline hyperbolically. This process has high persistence, or 'long memory', but is nonetheless mean-reverting for $v > 0$. The hyperbolic-decline property is shared by the fractionally integrated or ARFIMA(p, d, q) class of process, which take the form

$$x_t = (1-L)^{-d} u_t$$

Where u_t is a stationary ARMA(p, q) process, with $d = 1 - v$. The ARFIMA model, plus a possible deterministic component, is accordingly proposed as a plausible model to represent the time series of log[$\overline{X}_t/(1-\overline{X}_t)$]. When d is close to 1 the series is accordingly more persistent, as is expected since the parameter v is close to 0 when the distribution of α^i is concentrated near 1. The degree of persistence of the aggregate process therefore depends on the proportion of committed voters in the population.

3.0 Theory and Methods

3.1 The ARMA process

Modelling of stochastic time series generated by the ARMA or ARIMA processes can be done with the Box-Jenkins approach. The methods and procedures can be found in the literature [11].

3.1.1 Exact Maximum Likelihood Estimation of AR(1)

The covariance stationary first-order Gaussian autoregressive process is $x_t = \rho x_{t-1} + \varepsilon_t$ such that $\varepsilon_t \sim N(0, \delta^2)$ where $|\rho| < 1$, $t = 1, ..., T$. The likelihood may be factored into the product of $T-1$ conditional likelihoods and an initial marginal likelihood [12]. Specifically,

$$L(\phi) = l_T(y_T \mid \Omega_{T-1};\phi)\, l_{T-1}(y_{T-1} \mid \Omega_{T-2};\phi) \ldots l_2(y_2 \mid \Omega_1;\phi)\, l_1(y_1;\phi)$$

Where $\phi = (\rho, \sigma^2)'$ and $\Omega = \{y_t, ..., y_1\}$. The initial likelihood term $l_1(y_1;\phi)$ is known in closed form; it is

$$l_1(y_1;\phi) = (2\pi)^{-1/2}\sqrt{\frac{1-\rho^2}{\sigma^2}}\exp\left[-\frac{1-\rho^2}{2\sigma^2}y_1^2\right]$$

The remaining likelihood terms are

$$l_t(y_t \mid \Omega_{t-1};\phi) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y_t - \rho y_{t-1})^2\right] \qquad t = 2,...,T$$

Beach and MacKinnon [13] show that small-sample bias reduction and efficiency gains are achieved by maximising the exact likelihood, which includes the initial likelihood term, as opposed to the approximate likelihood, in which the initial likelihood term is either dropped or treated in an ad hoc manner. Moreover, they find that as ρ increases, the relative efficiency of exact maximum likelihood increases.

At any numerical iteration, say j^{th}, on the way to finding a maximum of the likelihood, a current "best guess" of the parameter vector exists; $\phi^{(j)}$.

The Gaussian likelihood is then constructed as

$$L(\phi^{(j)}) \approx \hat{l}_1(y_1,;\phi^{(j)}) \prod_{t=2}^{T} \sigma^{-1} \exp\left[-\frac{1}{2\sigma^2}(y_t - \rho y_{t-1})^2\right],$$

and it is maximised with respect to ϕ using standard numerical techniques

3.2 The ARFIMA process

The fractionally integrated ARMA model denoted ARFIMA(p,d,q) has become increasingly popular to describe time series that exhibit long [14]. Granger and Joyeux [10] and Hosking [15] introduced fractional differencing and the general class of autoregressive fractionally integrated moving average (ARFIMA) models. Let B denote the lag operator $BX_t = X_{t-1}$. Then the stationary and invertible ARFIMA(p,d,q) model is written as

$$\Phi(L)(1-L)^d(y_t - \mu) = \Theta(L)\varepsilon_t,\ \varepsilon_t \sim NID(0,\sigma_t^2) \qquad (3)$$

where d is the fractional integration parameter, $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ specifies the AR lag polynomial, and $\Theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$ the MA polynomial. For stationarity and invertibility, the roots of $\Phi(z)$ and $\Theta(z)$ must lie outside the unit circle. The integration parameter d can assume values based on two conditions ([16]; [17]; [18]; [19]): A series exhibits a stationary and invertible ARMA process with geometrically bounded autocorrelations if $-0.5 < d < 0.5$ ([15], [19]). Secondly, it exhibits non-stationary process if $0.5 \le d < 1$. For $0 < d < 0.5$, X_t is a stationary long memory process in the sense that autocorrelations are not absolutely summable and decay exponentially to zero [19].

3.3 Estimating ARFIMA models

Mills [20] generates autocorrelation recursively as $\rho_1 = \frac{d}{1-d}$, $\rho_2 = \frac{d(1+d)}{(1-d)(2-d)}$ and

$$\rho_k = \prod_{i=0}^{k}\left(\frac{i-1+d}{i-d}\right) = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(d)\Gamma(k-d+1)} \approx \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \approx mk^{2d-1}$$

The autocovariance function of a stationary ARMA process with mean μ,

$$\gamma_i = \mathrm{E}[(y_t - \mu)(y_{t-1} - \mu)]$$

defines the variance matrix of the joint distribution of $y = (y_1,..., y_T)'$:

$$V[y] = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \gamma_1 \\ \gamma_{T-1} & \cdots & \cdots & \gamma_0 \end{bmatrix} = \Sigma, \qquad (4)$$

which is a symmetric Toeplitz matrix, denoted by $\tau[\gamma_0,...,\gamma_{T-1}]$. Under normality:

$$y \sim N_T(\mu,\Sigma)$$

and combined with a procedure to compute the autocovariances in (4), the log likelihood (writing $z = y - \mu$) is given by

$$\log\ L(d,\phi,\theta,\beta,\sigma_\varepsilon^2) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}z'\Sigma^{-1}z\,. \qquad (5)$$

Additional regression parameters in μ are denoted by β, but can be ignored initially. The autocovariances of a stationary ARMA process scaled by the error variance, $r_i = \gamma_i/\sigma_\varepsilon^2$ is $R = \tau[\gamma_0/\sigma_\varepsilon^2,...,\gamma_{T-1}/\sigma_\varepsilon^2]$. In order to allow maximum likelihood estimation, the autocovariance function must be evaluated in order to construct the autocovariances, R. An algorithm for the computation of autocovariances of the ARFIMA process (3) is derived in ([16]:

$$\gamma_i = \sigma_\varepsilon^2 \sum_{k=-q}^{q}\ \sum_{j=1}^{p} \Psi_k \zeta_j C(d, p+k-i, \rho_j) \qquad (6)$$

where $\rho_1,...,\rho_p$ are the (possibly complex) roots of the AR polynomial and

$$\Psi_k = \sum_{s=|k|}^{q} \theta_s \theta_{s-|k|}\ , \qquad (7)$$

$$\zeta_j^{-1} = \rho_j \left[\prod_{i=1}^{p}(1-\rho_i\rho_j)\prod_{\substack{m=1\\ m\neq j}}^{p}(\rho_j - \rho_m)\right] \qquad (8)$$

where $\theta_0 = 1$. C is defined as

$$C(d,h,\rho) = \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2}\frac{(d)_h}{(1-d)_h} \times \left[\rho^{2p}F(d+h,1;1-d+h;\rho) + F(d-h;1;1-d-h;\rho) - 1\right] \qquad (9)$$

where Γ is the gamma function, ρ_j are the roots of the AR polynomial and $F(a,1;c;\rho)$ is the hypergeometric function

$$F(a,b;c;\rho) = \sum_{i=0}^{\infty}\frac{(a)_i(b)_i}{(c)_i}\frac{\rho^i}{i!} \qquad (10)$$

where we use Pochhammer's symbol

$$(a)_i = a(a+1)(a+2)...(a+i-1),\ (a)_0 = 1\,. \qquad (11)$$

Computation of $F(a,1;c;\rho)$ can be done recursively,

$$F(a,1;c;\rho) = \frac{c-1}{\rho(a-1)}\left[F(a-1,1;c-1;\rho) - 1\right] \qquad (12)$$

In the absence of AR parameters reduces to

$$\gamma_i = \sigma_\varepsilon^2 \sum_{k=-q}^{q} \Psi_k \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2}\frac{(d)_{k-i}}{(1-d)_{k-i}} \qquad (13)$$

the ratio $\frac{(d)_h}{(1-d)_h}$ for $h = p-q-T+1,...,0,..., p+q$ can be computed using a forward recursion for $h > 0$

$$(d)_h = (d+h-1)(d)_{h-1},\ h > 0, \qquad (14)$$

and a backward recursion otherwise:

$$(d)_h = \frac{(d)_{h-1}}{(d-h)},\ h < 0\,. \qquad (15)$$

3.4 Data

The political opinion polls data used were obtained from the Infotrak Harris Research, Consumer Insight Research and Strategic Research for the period between September and December 2007. Out of the 12 observations, 10 are used for model building while the remaining two observations are used to evaluate the forecast performance of the fitted models.

4.0 Results

The preliminary analysis of the data was done by use of time plots which provided basic characteristics of the series. The examination of the time plots in figures 1 to 3 shows that there is stationarity in the observed data. However, the most exact information on the memory decay process is obtained by estimating the decay rate, d.

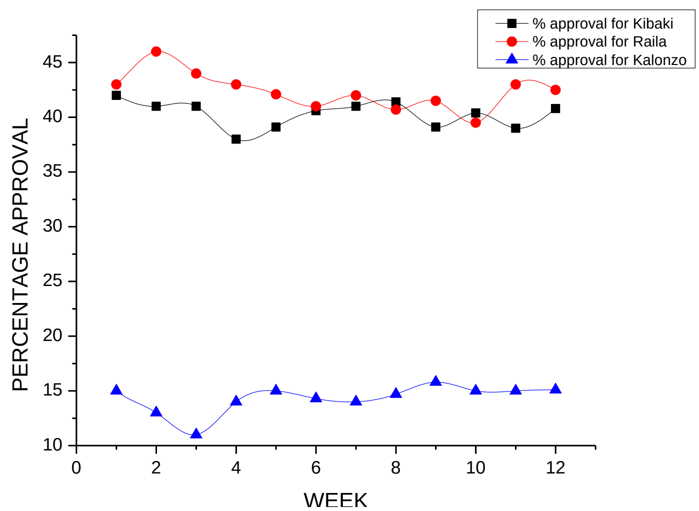

Figure 1: Time series plot for the Consumer Insight series

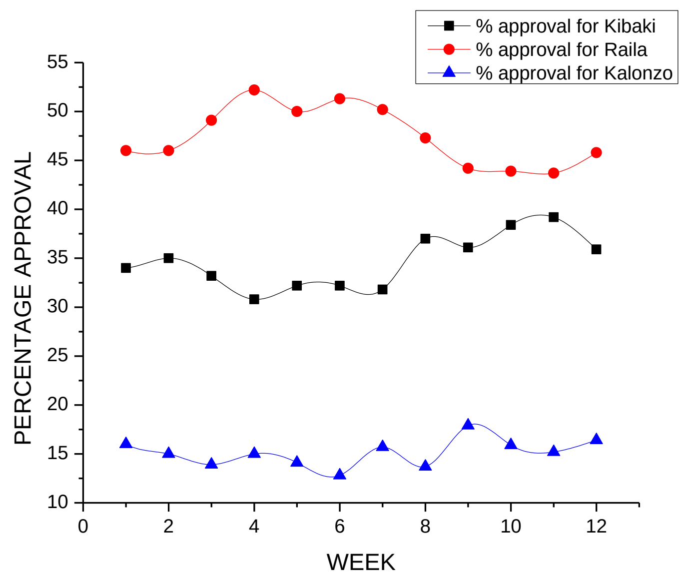

Figure 2: Time series plot for Infotrak Harris series

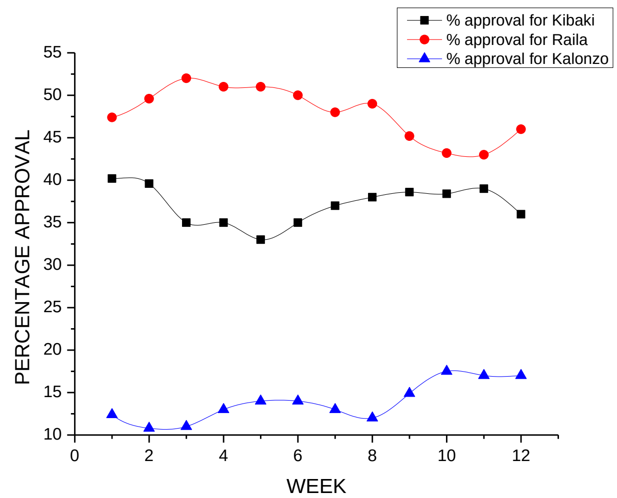

Figure 3: Time series plot for Strategic Research series

4.1 Unit root test

To apply the ARFIMA and ARIMA tests the series are examined for unit root and stationarity. Among the widely applied stationarity tests which include an option for fractional unit roots are the variance ratio test, Rescaled range test, Schmidt-Phillips test and the KPSS test [21]. However these tests share the severely limiting weakness that a long time series ($n \geq 1000$) is needed to distinguish long memory from short memory reliably. An adaptable stationarity test for a series with less than 200 observations is the ADF test. It does not directly indicate whether the series has a fractional unit root but this weakness can be covered if we can conclude that a series possibly has a fractional unit root when both alternatives are excluded.

The most exact information on the memory decay process, however, is obtained by estimating the decay rate, d . Fractional integration estimates also simplify the analysis of time series data by ending debates over the best way to test for unit roots, where one needs to choose among many different tests, such as Dickey Fuller, ADF, variance ratio, or KPSS, and by assumption choose the null hypotheses of d = 1 or d = 0, that is, instead of running multiple tests and looking for patterns suggesting stationarity, one can instead rely upon the point estimates of d [2]. There are three methods of doing this: semiparametric estimation [22], the approximate maximum likelihood in the frequency domain ([23], [24]) and the exact maximum likelihood in the time domain ([16]. The first two do not perform well in small samples ([16]. In this study the maximum likelihood method was used to estimate the decay rate d.

4.2 ARFIMA modelling of the polls data

Appropriate ARFIMA models are fitted to the polls data. The long memory parameter d is estimated using the maximum likelihood method. The estimated long memory parameter is used to fractionally difference the polls series which is then modelled as an ARFIMA process.

The corrected AIC is used as the model identification tool. The optimal model for the fractionally differenced series is the one that has the least value of the corrected AIC.

The exact maximum likelihood estimation technique is adopted to estimate the model parameters. The exact likelihood procedure appears more suitable than approximate procedures when working with small data sets and particularly when estimating models whose characteristic equations have roots close to the boundary of the unit circle[24]. Interest in the closed form of the likelihood function of the model stems from the need to make inferences in the following situations; when the sample is small, when the parameters are close to the invertibility boundary, and as an inference function for robust and missing observation problems. Sowell ([16] has also applied exact MLE technique to estimate the parameters of a univariate fractionally integrated time series. He also looked at the small sample properties of the estimators.

Optimal models for the fractionally differenced polls data together with the corrected AIC values and 95% confidence intervals for the parameter estimates are given in tables 1 – 3. Out of the nine series seven of them exhibit long memory characteristics with the value of the fractional differencing parameter ranging between 0.1 and 0.4. Six of models are pure fractional processes, ARFIMA(0, d, 0), while one is ARFIMA(0, d, 1) process. The estimates of the fractional differencing parameter d all fall within the 95% confidence interval. Of interest is the fact that for all ARFIMA models, d = 0 and d = 1 do not fall within the confidence intervals. Taking the first differences would therefore lead to overdifferencing. Therefore, in spite of the small sample sizes, we may conclude that the series are stationary and exhibit long-range dependence due to the fact that $d < 0.5$.

Table 1: Optimal ARFIMA(p, d, q) models for the Consumer Insight data

Series	Fitted models	95% CI for d	AICc
Raila series	ARFIMA(0, 0.3395859, 0)	0.3395855 - 0.3395863	42.66
Kalonzo series	ARFIMA(0, 0.09802818, 0)	0.09802783 - 0.09802854	38.96

Table 2: Optimal ARFIMA(p, d, q) models for the Infotrak Harris data

Series	Fitted models	95% CI for d	AICc
Kibaki series	ARFIMA(0, 0.2990022, 0)	0.2990017 - 0.2990027	49.71
Raila series	ARFIMA(0, 0.3528283, 0)	0.3528278 0.3528288	52.51

Table 3: Optimal ARFIMA(p, d, q) models for the Strategic Research data

Series	Fitted models	95% CI for d	Q	AICc
Kibaki series	ARFIMA(0, 0.3062491, 0)	0.3062487 0.3062496	-	49.21
Raila series	ARFIMA(0, 0.368705, 0)	0.3687045 0.3687055	-	50.4
Kalonzo series	ARFIMA(0, 0.3342121, 1)		-0.8346534	42.77

Standard diagnostic tests are done using the residual ACF and PACF of the models. The models pass these tests since their residual values lie within the 95% confidence interval band. Residual plots of the ARFIMA (p, d, q) models are also plotted to examine whether the models are white noise or not. The presence of residuals correlation is tested by use of the [26] test. The [27] test is used to test the normality of the residuals. These are tests for model adequacy and the results are reported in tables 4 - 6. The Ljung-Box test rejects the presence of serial correlation since all the p-values are greater than 0.05. The models are therefore adequate at the 5% level. The Jarque-Bera tests also show that the residuals are normally distributed since the p-values are all greater than the 0.05 level of significance.

Table 4: Model checking of ARFIMA(p, d, q) models for the Consumer Insight data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Raila series	ARFIMA(0, 0.3395859, 0)	Ljung-Box (lag 3)	1.7221	0.632
		Jarque-Bera	2.2411	0.3261
Kalonzo series	ARFIMA(0, 0.09802818, 0)	Ljung-Box (lag 3)	2.9545	0.3987
		Jarque-Bera	3.5051	0.1733

Table 5: Model checking of ARFIMA(p, d, q) models for the Infotrak Harris data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Kibaki series	ARFIMA(0, 0.2990022, 0)	Ljung-Box (lag 3)	2.4105	0.4917
		Jarque-Bera	0.6202	0.7334
Raila series	ARFIMA(0, 0.3528283, 0)	Ljung-Box (lag 3)	4.6635	0.1982
		Jarque-Bera	0.5212	0.7706

Table 6: Model checking of ARFIMA(p, d, q) models for the Strategic Research data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Kibaki series	ARFIMA(0, 0.3062491, 0)	Ljung-Box (lag 3)	4.7183	0.1936
		Jarque-Bera	0.847	0.6548
Raila series	ARFIMA(0, 0.368705, 0)	Ljung-Box (lag 3)	3.4343	0.3294
		Jarque-Bera	0.5154	0.7728
Kalonzo series	ARFIMA(0, 0.3342121, 1)	Ljung-Box (lag 3)	3.807	0.2831
		Jarque-Bera	0.6192	0.7338

4.3 Forecasting evaluation of the fitted ARFIMA models

After model checking, the forecast values are studied. Both the in-sample and out-of-sample values are computed. The forecast performance of the models is evaluated by use of the Root Mean Square Error (RMSE). The model with the lowest value of these forecast evaluation tools is considered to have the best prediction power. The out-of-sample forecast values together with RMSE values are shown in tables 7 – 13.

Table 7: Out-of-sample forecasts for the ARFIMA(0, 0.3395859, 0) model fitted to Raila series from Consumer insight

Week	Optimal forecast	Actual	Error
11	41.19523	43	1.80477
12	51.52955	42.5	0.97045
RMSE	0.647994		

Table 8: Out-of-sample forecasts for the ARFIMA(0, 0.09802818, 0) model fitted to Kalonzo series from Consumer insight

Week	Optimal forecast	Actual	Error
11	14.31977	15	0.68023
12	14.26231	15.1	0.83761
RMSE	0.341219		

Table 9: Out-of-sample forecasts for the ARFIMA(0, 0.2990022, 0) model fitted to Kibaki series from Infotrak Harris

Week	Optimal forecast	Actual	Error
11	35.49741	39.2	3.70259
12	34.98894	35.9	0.91106
RMSE	1.205786		

Table 10: Out-of-sample forecasts for the ARFIMA(0, 0.3528283, 0) model fitted to Raila series from Infotrak Harris

Week	Optimal forecast	Actual	Error
11	46.36690	43.7	-2.6669
12	46.91363	45.8	-1.11363
RMSE	0.913922		

Table 11: Out-of-sample forecasts for the ARFIMA(0, 0.3062491, 0) model fitted to Kibaki series from Strategic Research

Week	Optimal forecast	Actual	Error
11	37.50152	39	1.49848
12	37.31383	36	-1.31383
RMSE	0.630206		

Table 12: Out-of-sample forecasts for the ARFIMA(0, 0.368705, 0) model fitted to Raila series from Strategic Research

Week	Optimal forecast	Actual	Error
11	46.42137	43	-3.42137
12	47.12248	46	-1.12248
RMSE	1.138672		

Table 13: Out-of-sample forecasts for the ARFIMA(0, 0.3342121, 1) model fitted to Kalonzo series from Strategic Research

Week	Optimal forecast	Actual	Error
11	16.16380	17	0.8362
12	14.70307	17	2.29693
RMSE	0.772989		

4.4 ARMA modelling of the polls data

Appropriate ARMA models are fitted to the Kenyan presidential approval polls data. Model selection is done by use of the corrected AIC. After the best models are chosen, the parameters of the models are examined next. The parameters are estimated by the exact maximum likelihood estimation method which is known to perform better in small samples. The results of the parameter estimates of the optimal models together with the values of the corrected AIC are shown in tables 14 – 16.

Table 14: Optimal ARIMA(p, d, q) models for the Consumer Insight data

Series	Fitted models	p	q	AICc
Kibaki series	ARMA(0, 1)	-	0.2154	41.23
Raila series	ARMA(1, 0)	0.5775598	-	46.12
Kalonzo series	ARMA(0, 1)	-	0.9819	39.46

Table 15: Optimal ARIMA(p, d, q) models for the Infotrak Harris data

Series	Fitted models	p	q	AICc
Kibaki series	ARMA(1, 0)	0.5552499	-	52.7
Raila series	ARMA(0, 1)	-	0.7345	53.24
Kalonzo series	ARMA(1, 0)	-0.084131	-	44.86

Table 16: Optimal ARIMA(p, d, q) models for the Strategic Research data

Series	Fitted models	p	q	AICc
Kibaki series	ARMA(1,0)	0.6202688	-	50.25
Raila series	ARMA(0, 1)	-	0.9930	51.19
Kalonzo series	ARMA(0, 1)	-	1.0000	44.06

After the optimal models have been fitted, we next check for the adequacy of the models. The model adequacy is tested using two diagnostic tests, Ljung-Box test and Jarque-Bera test. Ljung-Box test is used to check for serial correlation in the residuals while the Jarque-Bera test is used to check for normality. As shown in tables 17 – 19, the residuals appear to be white noise and the data are normal since the p-values for the two diagnostic tests are all greater than 0.05 level of significance.

Table 17: Model checking of ARMA(p, q) models for the Consumer Insight data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Kibaki series	ARMA(0, 1)	Ljung-Box (lag 3)	4.2491	0.2358
		Jarque-Bera	0.9109	0.6342
Raila series	ARMA(1, 0)	Ljung-Box (lag 3)	0.7100	0.7010
		Jarque-Bera	0.5109	0.7746
Kalonzo series	ARMA(0, 1)	Ljung-Box (lag 3)	0.6298	0.8896
		Jarque-Bera	3.3988	0.1828

Table 18: Model checking of ARMA(p, q) models for the Infotrak Harris data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Kibaki series	ARMA(1, 0)	Ljung-Box (lag 3)	0.9000	0.6373
		Jarque-Bera	0.7514	0.6868
Raila series	ARMA(0, 1)	Ljung-Box (lag 3)	2.198	0.5323
		Jarque-Bera	0.7986	0.6708
Kalonzo series	ARMA(1, 0)	Ljung-Box (lag 3)	0.5100	0.7747
		Jarque-Bera	0.3434	0.8422

Table 19: Model checking of ARMA(p, q) models for the Strategic Research data

Series	Fitted models	Diagnostic test	Test statistic	p-value
Kibaki series	ARMA(1, 0)	Ljung-Box (lag 3)	1.6800	0.4324
		Jarque-Bera	0.6758	0.7133
Raila series	ARMA(0, 1)	Ljung-Box (lag 3)	1.1723	0.7597
		Jarque-Bera	1.0311	0.5972
Kalonzo series	ARMA(0, 1)	Ljung-Box (lag 3)	3.3816	0.3364
		Jarque-Bera	1.0428	0.5937

4.5 Forecasting evaluation of the fitted ARMA models

The out-of-sample forecast values of the fitted models are shown in tables 20 – 28 below. The Root Mean Square Error values are also displayed in the tables. RMSE values are used to evaluate the forecast performance of the fitted models.

Table 20: Out-of-sample forecasts for the ARMA(0, 1) model fitted to Kibaki series from Consumer Insight

Week	Optimal forecast	Actual	Error
11	40.46827	39	-1.46827
12	40.39893	40.8	0.40107
RMSE	0.481318		

Table 21: Out-of-sample forecasts for the ARMA(1, 0) model fitted to Raila series from Consumer Insight

Week	Optimal forecast	Actual	Error
11	40.67438	43	2.32562
12	41.35266	42.5	1.14734
RMSE	0.820055		

Table 22: Out-of-sample forecasts for the ARMA(0, 1) model fitted to Kalonzo series from Consumer Insight

Week	Optimal forecast	Actual	Error
11	14.29901	15	0.70099
12	14.18340	15.1	0.9166
RMSE	0.364903		

Table 23: Out-of-sample forecasts for the ARMA(1, 0) model fitted to Kibaki series from Infotrak Harris

Week	Optimal forecast	Actual	Error
11	36.47423	39.2	2.72577
12	35.40495	35.9	0.49505
RMSE	0.876065		

Table 24: Out-of-sample forecasts for the ARMA(0, 1) model fitted to Raila series from Infotrak Harris

Week	Optimal forecast	Actual	Error
11	46.72746	43.7	-3.02746
12	47.81429	45.8	-2.01429
RMSE	1.149908		

Table 25: Out-of-sample forecasts for the ARMA(1, 0) model fitted to Kalonzo series from Infotrak Harris

Week	Optimal forecast	Actual	Error
11	14.92428	15.2	0.27572
12	15.00637	16.4	1.39363
RMSE	0.449247		

Table 26: Out-of-sample forecasts for the ARMA(1, 0) model fitted to Kibaki series from Strategic Research

Week	Optimal forecast	Actual	Error
11	37.86078	39	1.13922
12	37.52632	36	-1.52632
RMSE	0.602285		

Table 27: Out-of-sample forecasts for the ARMA(0, 1) model fitted to Raila series from Strategic Research

Week	Optimal forecast	Actual	Error
11	46.95964	43	-3.95964
12	48.42653	46	-2.42653
RMSE	1.468564		

Table 28: Out-of-sample forecasts for the ARMA(0, 1) model fitted to Kalonzo series from Strategic Research

Week	Optimal forecast	Actual	Error
11	15.34	17	1.66
12	13.44	17	3.56
RMSE	1.242143		

The observed and forecasted values are closer for the ARFIMA models than for the ARMA models. This shows that ARFIMA models forecast better the given data than the ARMA models. Besides, the RMSE values for the ARFIMA models are found to be lower than those of the ARMA models in five of the seven compared models. ARMA(1, 0) models seem to be forecasting better than ARFIMA(0, d, 0) models for the Kibaki series obtained from Infotrak Harris and Strategic Research. Therefore, ARFIMA models are generally found to be the better models for the Kenyan presidential approval data than the ARMA models.

5.0 Conclusion

Long memory ARFIMA(p, d, q) models are used to fit the Kenyan presidential approval data. Seven out of the nine series are found to exhibit stationary long-range dependence with the estimates of the memory decay parameter d ranging between 0.1 and 0.4. Six of these models are pure fractional models while one has the short-memory Moving Average component. AR(1) and MA(1) models are also fitted to the same data. Even though all the models fit the data well, the forecasts obtained using the long memory models resemble the actual values better than the forecasts using the short memory models in five of the seven models compared.

In popularity series the assumption of stationarity means very stable popularity shares because of mean reversion [6]. Gelman and King [28] make an argument about campaigns activating people's 'enlightened preferences'. They argue that campaigns matter because they inform potential voters, and as potential voters become more informed, their preferences begin to change. They argue that if the series is Nonstationary it indicates that the public's presidential preferences during the general election campaign did not simply bounce around a constant mean but rather trended somewhere.

It is not, as the election forecasting perspective might suggest, that voters knew the final answer right from the start, but instead voters underwent a process whereby they eventually reached the final answer in the end. If the series was stationary, it would suggest that voter's preferences really did not move much during the general election. ARFIMA models for Kenya's polls data fall within the stationary regime which suggests that Kenyan voters seem to have a prior choice of their preferred presidential candidate and even the campaigns do not change very much their opinions. These fractional differencing parameter values are less than those found in mature democracies but close to those found in smaller regional parties in Spain [6]. The lower values of long memory parameter imply the existence of some significant differences in the behaviour of Kenyan voters and those voters in more mature democracies. The implication is that the candidates had a small share on 'non-militant' supporters with the 'militant' or die-hard supporters taking the most prominence.

To our knowledge, ARFIMA and ARMA models have not been used to model opinion poll data in Kenya. This is therefore one of the contributions of this research. The second contribution of the research is to be able to predict the outcome of the elections.

We suggest further research on the Kenya's polls data for the influence of such factors as election cycle, structural break and the regional support influence. Kenyan opinion polls have not been tested before for fractional integration. Therefore, there is no comparable evidence for these results. This leaves the possibility of confusion between long/perfect memory and structural break still open.

References

Box-Steffensmeier, J.M. and Smith, R.M.,The dynamics of aggregate partisanship. *American political science review*, 90, 567-580 (1996)

Box-Steffensmeier, J.M. and Smith, R.M.. Investigating political dynamics using fractional integration methods. *American journal of political science*, 42(2), 661-689 (1998)

Byers, J.D., Davidson, J. and Peel, D.A., Modelling political popularity: An analysis of long range dependence in opinion poll series. *Journal of the Royal statistical society*. Series A, 160, 471-490 (1997)

Byers, J.D. Davidson, J. and. Peel, D.A., The dynamics of aggregate political popularity: evidence from eight countries. *Electoral Studies*, 19, 49-62 (2000)

Eisinga, R., Franses, P.H. and Ooms, M., Forecasting long memory left-right political orientations. *International journal of forecasting*, 15, 185-199 (1999)

Dolado, J.J., Gonzalo, J. and Mayoral, L., Long range dependence in Spanish political opinion poll series. *Journal of applied econometrics.* 18(2), 137-155 (2003)

Lebo, M.J., Walker, R.W. and Clarke, H.D., You must remember this: Dealing with long memory in political analysis. *Electoral Studies*, 19, 31-48 (2000)

Clarke, H.D. and Lebo, M., Fractional (Co-integration) and governing party support in Britain.*British journal of political science*, 33, 283-301 (2003)

Franklin, C. H. and Jackson, J.E., The Dynamics of Party Identification. *American Political Science Review* 77: 957-73 (1983)

Granger, C.W.J. and Joyeux, R., An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1, 15-29 (1980)

Box, G.E.P. and Jenkins, G.M., *Time Series Analysis Forecasting and Control*. San Francisco: Holden. Day (1976)

Diebold, F. X. and T. Schuermann., Exact Maximum Likelihood Estimation of Observation-Driven Econometric Models," in R. S. Mariano, M. Weeks and T. Schuermann, eds., *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge: Cambridge University Press (1998)

Beach, C.M. and MacKinnon, J.G., A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica*, 46, 51-58 (1978)

Lieberman, O., Rousseau, J. and Zucker, D.M., Small-sample likelihood-based inference in the ARFIMA model. *Economic theory*, 16, 231-248 (2000)

Hosking, J.R.M., Fractional differencing. *Biometrika*, 68, 165-76 (1981)

Sowell, F.B., Maximum likelihood estimation of stationary univariate Fractionally Integrated time series models. *Journal of econometrics*, 53, 165-188 (1992)

Hurvich, C.M. and Ray, B.K., Estimation of the memory parameter by non-stationary and non-invertible fractionally integrated processes. *Journal of time series analysis,* 16, 17-41 (1995)

Chong, T.T.L.,The polynomial aggregated AR(1) model. *Econometrics journal*, 9, 98-122 (2006)

Mayoral, L., Minimum distance estimation of stationary and non-stationary ARFIMA processes. *Journal of econometrics,* 10, 124-148 (2006)

Mills, T.C.,Time series modelling of two millennia of Northern Hemisphere temperatures: long memory or shifting trend? *Journal of Royal statistical society,* 170(a), 83-94. processes. *Journal of econometrics,* 10, 124-148(2006)

Shittu, O.I. and Yaya, O.S.,Measuring forecast performance of ARMA and ARFIMA models: an application to US Dollar/UK Pound foreign exchange rate. *European journal of scientific research,* 32, 167-176 (2009)

Geweke, J. and Porter-Hudak, S.,The estimation and application of long-memory time series models. *Journal of time series analysis*, 4, 221-38 (1983)

Li, W.K. and A.I. McLeod, Fractional time series modelling, Biometrika 73, 217-221 (1986)

Fox, R. and Taqqu, M.S., Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of statistics,* 14, 517-532 (1986)

Nicholls, D.F. and Hall, A.D., The exact likelihood function of multivariate autoregressives moving average models. *Biometrika*, 66, 259-264 (1979)

Ljung, G.M. and Box, G.E.P., On a measure of lack of fit in time series models. *Biometrika,* 65, 297-303 (1978)

Jarque, C.M. and Bera, A.K., A test of normality of observations and regression residuals. *Int.Stat. Rev.*, 55, 163-172 (1987)

Gelman A. and King G., "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science,* 23: 409-451 (2003)