Performance Comparison of Various Forecasting Techniques

ARIMA vs ARIMA using Error Correction and Ensemble Model

Surjit Laha
surjit.laha@bridgei2i.com

Krishna Mohan Roy
krishna.roy@bridgei2i.com

Abstract

This paper explores the possibility of improvement in pure ARIMA model through introduction of error correction adjustments in the context of forecasting business metrics like number of churns. It also provides detailed comparison of performance metrics of ARIMA model and an ensemble model that uses error corrected forecast of ARIMA, exponential smoothing (ES) and moving average. In the real life experimentation, mean absolute percentage error (MAPE) obtained from individual models are used as weightage of ensemble model.

Keywords: ARIMA, ES, TS, MAPE, Ensemble are used as weightage of ensemble model.

1. Introduction

Attrition of customers in any industry is a major concern for companies. Deactivation, also called churn causes revenue loss to the company as well as it causes problem in resource allocation and planning. If extent of deactivation is known in advance then company can plan resource allocation. This way they can save unnecessary spend and also plan to provide better satisfaction to their customer. There are many ways in which future deactivation can be estimated like Delphi method or some statistical method.

Forecasting techniques like Delphi method is judgemental forecast whereas methods like ARIMA and Exponential Smoothing (ES) are statistical method. ARIMA and ES are two widely used methods for forecast. ARIMA and ES method can handle different type of time series data which include trend, seasonality and level. Naïve based forecast where last observation is considered as forecast and forecast based on simple average of given number of past observation are also famous in some industry domain like finance.

This paper will leverage real life use case of telecom churn as an example. A time series forecast model was built to project number of deactivations expected for a US based Fortune 1000 Telecom Company using market level deactivation data available at monthly frequency. Deactivation was done for all markets at month level. We have forecasted at same periodicity. Deactivations for 20+ markets were forecasted based on data available over 3+ years.

This paper contains six sections. Section 2 explains various approach for time series forecast. Section 3 contains methodology followed for the proposed model. Section 4 contains comparative analysis of results obtained from the proposed model and ARIMA, it also compares performance of ARIMA with proposed model. Section 5 concludes this paper.

2. Literature Survey

2.1 De-seasonalizing Data

If there is seasonality in data and seasonality is very stable then we can de-seaosnalize data. Seasonality index of given month (taking 24 months data) can be given by Eq. (1)

$$SI_i = \frac{(Y_i + Y_{i+12})/2}{\frac{1}{24} * \sum_{j=1}^{24} Y_j} \qquad (1)$$

Let's say Yi be observation for the ith month of the year, then de-seasonalized value for the ith month, Y_i^d is given by Eq. (2)

$$Y_i^d = \frac{Y_i}{SI_i} \qquad (2)$$

2.2 Moving Average

In most of the cases like finance naïve base method or simple average of last n (n=3, 6 here) observations work quite well. We have taken average of last 6 observations. Let's assume we have data till time T and last six observations are Y_T, Y_{T-1}, Y_{T-2}, Y_{T-3}, Y_{T-4} and Y_{T-5} then forecast for next observation using moving average is given by Eq. (3)

$$\tilde{Y}_{T+1|T} = \frac{\sum_{i=0}^{n-1} Y_{T-i}}{n} \qquad (3)$$

Forecast for the next month (offset two) is calculated using latest n-1 number of observations and offset one forecast. $\tilde{Y}_{T+2|T}$ is given by Eq. (4)

$$\tilde{Y}_{T+2|T} = \frac{\tilde{Y}_{T+1|T} + \sum_{i=1}^{n-1} Y_{T-i}}{n} \qquad (4)$$

Moving average with window size n, is obtained from Eq. (3). By taking n=3, 6 we can calculate MA3 and MA6 respectively.

2.3 Exponential smoothing

This class of forecast model was developed by Holt, Brown and Winter's in late 1950s. ES method uses weighted average of past data to arrive at forecast future. Weight used to average decay exponentially i.e. older data points get less weight compared to present data points. This method assume that the underlying time series have level, trend and seasonality, so ES find these component in the time series.

There are mainly three types of ES, which are as follows:

1. Simple Exponential Smoothing: when time series data don't have any trend or seasonality and have only level then this method could be used to forecast. This is the simplest of all ES methods [1] [2].

$$\tilde{Y}_{T+1|T} = \alpha Y_T + \alpha(1-\alpha)Y_{T-1} + \alpha(1-\alpha)^2 Y_{T-2} + \cdots \qquad (5)$$

Where,

α is a smoothing parameter and $0 \le \alpha \le 1$

$\tilde{Y}_{T+1|T}$ is the forecast at time T+1 using data till time T

Y_T is the actual value at time T

$\tilde{Y}_{T+1|T} = l_T$ Forecast equation   (6)

$l_T = \alpha Y_T + (1-\alpha)l_{T-1}$ Level equation   (7)

Where,

l_T is level at time T

2. Holt's Linear Trend Method: when time series data have level and trend and no seasonality then this method could be used to forecast. This method has two component, level and trend [1] [2]

$$\tilde{Y}_{T+n|T} = l_T + nb_T \text{ Forecast equation} \qquad (8)$$
$$l_T = \alpha Y_T + (1-\alpha)(l_{T-1} + b_{T-1}) \text{ Level equation} \qquad (9)$$
$$b_T = \beta(l_T + l_{T-1}) + (1-\beta)b_{T-1} \text{ Trend equation} \qquad (10)$$

Where,

α and β are smoothing parameter and $0 \leq \alpha, \beta \leq 1$

$\tilde{Y}_{T+n|T}$ is the forecast at time T + n using data till time T

l_T is level at time T

b_T is trend at time T

3. Holt-Winters Seasonal Method: when time series data have level, trend and seasonality then this method could be used to forecast. This method has three component, level, trend and seasonality [1] [2]

$$\tilde{Y}_{T+n|T} = l_T + nb_T + s_{T-m+n_m} \tag{11}$$
$$n_m = floor(n-1) \; mod \; m + 1 \tag{12}$$
$$l_T = \alpha(Y_T - s_{T-m}) + (1 - \alpha)(l_{T-1} + b_{T-1}) \tag{13}$$
$$b_T = \beta(l_T + l_{T-1}) + (1 - \beta)b_{T-1} \tag{14}$$
$$s_T = \gamma(Y_T - l_{T-1} - b_{T-1}) + (1 - \gamma)s_{T-m} \tag{15}$$

Where,

α, β and γ are smoothing parameter and $0 \leq \alpha, \beta, \gamma \leq 1$

$\tilde{Y}_{T+n|T}$ is the forecast at time T + n using data till time T

l_T is level at time T

b_T is trend at time T

s_T is seasonality at time T

2.4 Autoregressive integrated moving average (ARIMA)

ARIMA models aim to describe the autocorrelations in the data. This class of time series model are very famous for forecasting time series which are stationary and non-stationary (Non-stationary can be made stationary by difference) [1] [3].

$$\acute{Y}_T = c + \varphi_1 \acute{Y}_{T-1} + \ldots + \varphi_p \acute{Y}_{T-p} + \theta_1 e_{T-1} + \cdots + \theta_q e_{T-q} \tag{16}$$

Where,

c is some constant

$\varphi_1, \ldots, \varphi_p$ are AR coefficient

$\acute{Y}_{T-1}, \ldots, \acute{Y}_{T-p}$ are d times differenced time series with lag

$\theta_1, \ldots, \theta_q$ are MA coefficient

$e_{T-1}, \ldots, e_{T-q}$ are error terms

The above equation is ARIMA (p, d, q) which means that there are p numbers of AR (Autoregressive) component, d number of difference of time series and q number of MA (Moving average) component. One has to identify p, d, and q of the ARIMA model then find the corresponding coefficient.

One could also include predictor in ARIMA models, assume churn is also affected by new activation then activation could be used to forecast churn in ARIMA model.

2.5 Error Correction using Tracking Signal

Residual obtained from the model should be randomly distributed and it should not show any correlation with Xs. Non-random distribution means there is information left in the residuals which could not be extracted in the model. An unbiased model should also have residuals which varies in both side of zero. If residuals tend to remain in one side of zero continuously, it means there is bias in the model.

Using tracking signal (TS) we can determine if there is any bias in the model or not and if any can be corrected [4-5]. Tracking signal is given by Eq. (17)

$$TS \quad = \quad \frac{\sum_{i=1}^{n} e_i}{MAD} \qquad (17)$$

Where e_i is residual (actual-forecast), n is period of forecast we take and MAD is given by Eq. (18)

$$MAD \quad = \quad \frac{\sum_{i=1}^{n} |e_i|}{n} \qquad (18)$$

Generally TS value anywhere between +4 and -4 means model is working well and we don't need any correction.

2.6 Ensemble Method

Given the nature of market behaviour, there may be the case when ARIMA performs better in one market while ES or moving average may perform better in other market. To make model robust we can use ensemble of all the models to get single model [6].

Ensemble model is nothing but weighted average of different models where weights are inverse of error they make in out of time data. More the error they make less is the weight they are given in final model.

Let's say ARIMA, ES and moving average make absolute percentage error of AE, EE and ME respectively. Forecast of ARIMA, ES and moving average are AF, EF and MF respectively then ensemble forecast will be given by Eq. (19)

$$Ensemble\ Forecast\quad =\quad \frac{\left(\frac{1}{AE}\right)*AF+\left(\frac{1}{EE}\right)*EF+\left(\frac{1}{ME}\right)*MF}{\left(\frac{1}{AE}+\frac{1}{EE}+\frac{1}{ME}\right)} \qquad (19)$$

We have calculated correction required from TS and using that correction factor we can correct next $\tilde{Y}_{T+1|T}$. Here, percentage error of ARIMA, AE can be given by Eq. (20)

$$AE\quad =\quad \frac{Corrected\ Arima\ Forecast-Actual}{Actual} \qquad (20)$$

Similarly, EE and ME can be calculated.

3. Methodology Used

For calculation of TS we have used offset one forecast for three consecutive months. Also we have considered +3 and -3 as TS cut off. If TS (17) lies outside the range [-3, 3] we calculate correction required as average percentage error, otherwise correction required is zero.

Correction calculated using TS is then applied to the offset one forecast of fourth month and error in corrected forecast from actual observation is calculated. Percentage error observed for each models work as their weight in ensemble model (19).

Our algorithm works in two stage, Stage I, where we choose model, find correction factor and weight for ensemble, whereas in Stage II, we do final forecast for offset one.

Figure 1 and Figure 2 show flow chart of algorithm. Below is the pseudo code for algorithm implementation.

Stage – I:

i. De-Seasonalize (1-2) the time series data for the market.

ii. Do one month offset forecast for 4 consecutive months for which you have actuals.

a. Forecast using ARIMA: make the time series stationary using differencing, if required. Say d is the difference required, value of d = [0, 2]. Calculating the AICc of different ARIMA models. Different ARIMA models formed are (p, d, q) with/without mean and with/without drift where p = [0, 2] and q = [0, 2]. Chose model with lowest AICc value and make forecast using it.

b. Forecast using ES: use Holt's linear additive forecast (11-15) method to forecast.

c. Forecast using MA6 and MA3: use MA6 and MA3 (3) to forecast.

So, we obtained 4 forecast numbers for each of the 4 consecutive months using each forecast model. Hence we have 4 forecast numbers for every month.

iii. Seasonalize all the forecast numbers from previous step.

iv. Calculate the TS (17) and percentage error correction using 1st, 2nd and 3rd month forecast and actual for all 4 forecast model. Correct the 4th month forecast number for all 4 forecast model. Calculate the TS and percentage error correction using 2nd, 3rd and 4th month forecast and actual for all 4 forecast model and save it to be used in stage II.

v. Calculate the weight for all 4 model using (20) which uses 4th month corrected forecast and actuals.

vi. Repeat step i to step v for all market

Figure 1. Schematic diagram of Forecasting Method
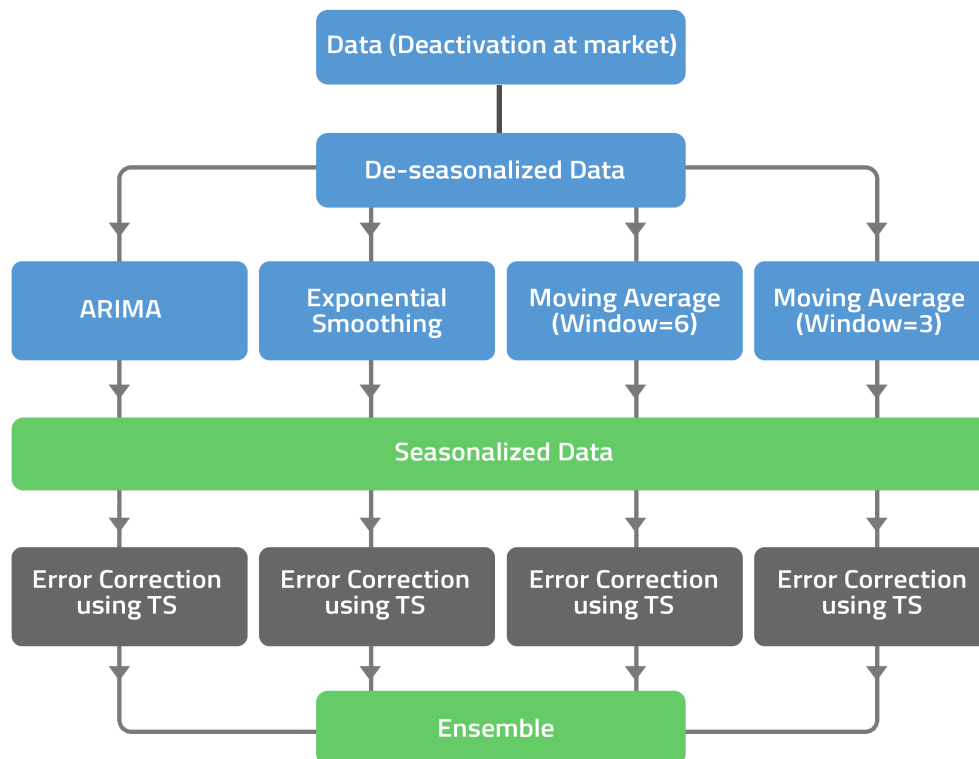
Weightage mechanism: Last month accuracy is used as weight to ensemble different models

Stage – II:

i. Forecast using all 4 forecast models for the desired months. Correct the forecast using percentage error correction calculated in Stage I step vi for all 4 forecast model.

ii. Ensemble all the forecast form 4 model using weight calculated from Stage I step v

iii. Repeat step i to step ii for all market.

Figure 2. Error Correction Algorithm
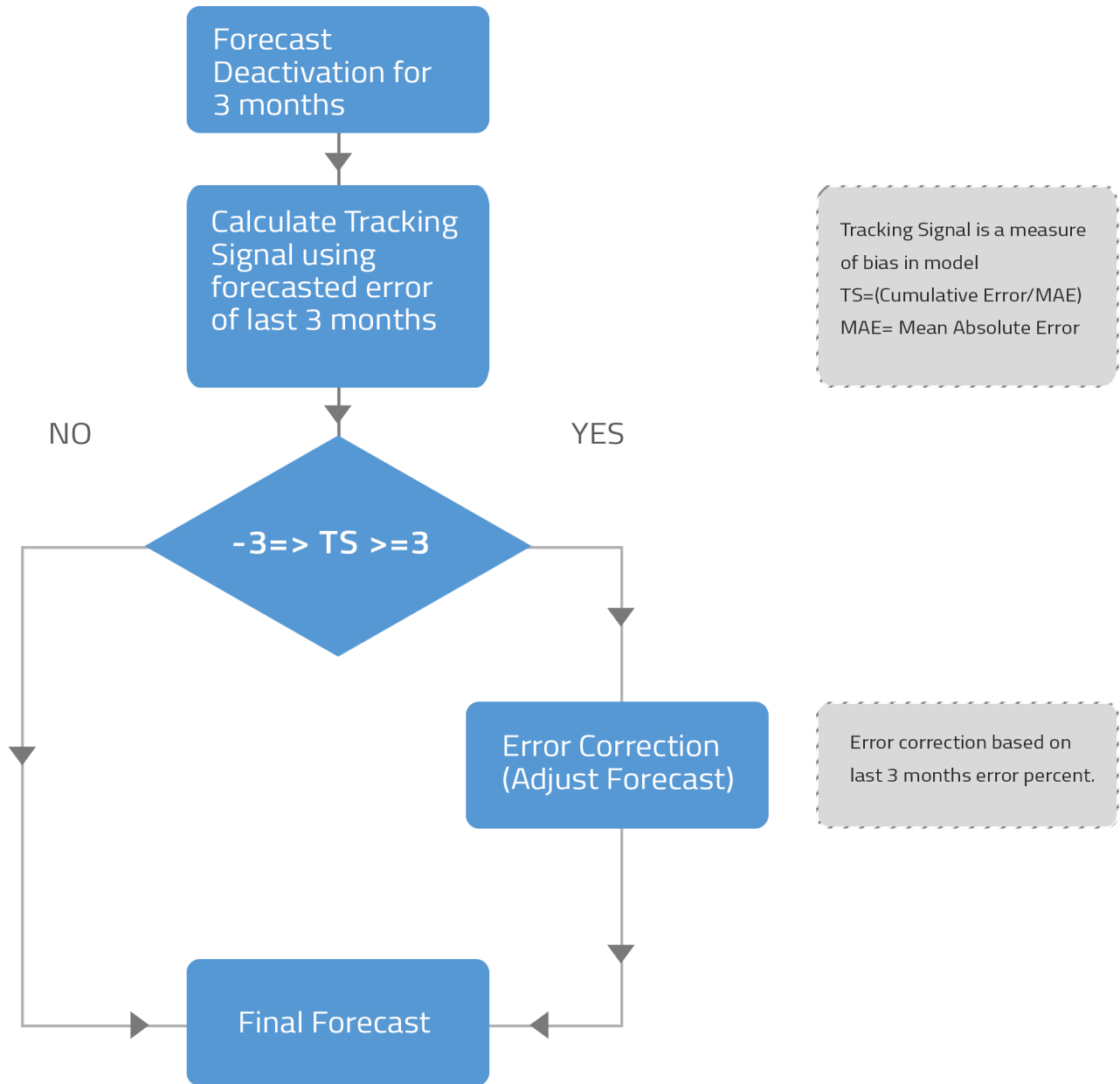
4. Results

Table 1-4 shows result for forecast for deactivation for 24 markets for 8 months.

4.1 ARIMA Vs ARIMA using Error Correction

Table 1. ARIMA Vs ARIMA using Error Correction

Methods	# Markets It is winner	%ERROR reduced per market
ARIMA	4	3.2%
Both equal	1	0%
ARIMA using Error Correction	19	4.9%

Table 2. Paired T-Test ARIMA Vs ARIMA using Error Correction

Mean Difference	3.32%
Standard deviation of difference	4.47%
Standard error of difference	0.93%
Lower confidence level	1.39%
Upper confidence level	5.25%
P-value two tail paired t-test	1.39E-03

As shown in Table 1, ARIMA using error correction performs better in 19 markets compared to ARIMA. On an average improvement per market is 4.9%.

Null hypothesis for the t-test:
H0: difference of mean of error of ARIMA and ARIMA using error correction is zero
H1: difference of mean of error of ARIMA and ARIMA with ensemble is not zero

Since p-value as shown in Table 2 is very small so null hypothesis, H0 can be rejected, also mean difference of MAPE is positive and so ARIMA with error correction performs better compared to ARIMA. At 95% confidence ARIMA using Error Correction performs better than ARIMA as shown in table 2.

4.2 ARIMA Vs Ensemble

As shown in Table 3, ensemble performs better in 22 markets compared to ARIMA model. On an average improvement per market is 5.4%.

Table 3. ARIMA Vs Ensemble

Methods	# Markets It is Winner	% Error Reduced per Market
ARIMA	2 Market	1.5%
Both Equal	0 Market	0%
Ensemble Forecast	22 Market	5.4%

Table 4. Paired T-Test ARIMA Vs Ensemble

Mean Difference	4.84%
Standard Deviation of Difference	4.59%
Standard Error of Difference	0.96%
Lower Confidence Level	2.86%
Upper Confidence Level	6.82%
P value two-tail paired t-test	3E-05

Null hypothesis for the t-test:

H_0: difference of mean of error of ARIMA and Ensemble is zero
H_1: difference of mean of error of ARIMA and ARIMA with ensemble is not zero

Since p-value as shown in Table 2 is very small so null hypothesis, H0 can be rejected, also mean difference of MAPE is positive and so Ensemble performs better compared to ARIMA. From table 4 we can easily observe that on 95% confidence Ensemble performs better than ARIMA.

5. Conclusion

From section 4.1 and 4.2 we can easily observe that ARIMA using error correction and ARIMA using ensemble performs better compared to ARIMA. Ensemble method also guarantee better performance from individual model performance.

6. References

[1]. Hyndman, R.J., Athanasopoulos, George (2013), Forecasting Principles and Practice, OTexts
[2]. Chatfield, C (2000), Time-series forecasting, Boca Raton: Chapman & Hall/CRC
[3]. Hyndman, R.J., A. B. Koehler, J. K. Ord and R. D. Snyder (2008), Forecasting with exponential smoothing: the state space approach, Berlin -Springer-Verlag
[4]. Gardner (1983), Automatic Monitoring of Forecast Errors, Journal of Forecasting, 18-22
[5]. Trigg, D (1964), Monitoring a Forecasting system, Operational Research Quarterly, 2718-74
[6]. Yaohui, Bai, Jiancheng Sun, Jianguo Luo, Xiaobin Zhang (6-8 Dec 2010), Forecasting financial time series with ensemble learning, in Intelligent Signal Processing and Communication Systems (ISPACS), pp.1-4, doi: 10.1109/ISPACS.2010.5704751

BRIDGE i2i
INFORMATION → INSIGHT → IMPACT

Follow Us