

**How does the latest Prophet
Forecasting model compare to the
exponential state space model
forecasts? Evaluation of
Bitcoin/ZAR predictability through
a Mincer-Zarnowitz approach**

Project Proposal

6 May 2017

Fiona Ganie (GNXFIO001)

GNXFIO001@myuct.ac.za

1 Introduction

Forecasting is a core element of the processes of a business. Producing forecasts of high quality is essential as they play an important role in inventory control, purchasing decisions of new equipment and investment decisions, and are needed for effective planning (Granvik 2010). Automatic forecasting techniques have commonly been used in business to produce a wide variety of forecasts and are a useful tool for producing a large number of forecasts. Automatic techniques can be used by non-experts without training in the use of time series (Hyndman and Khandakar 2008). Automatic forecasting algorithms must have the ability to select the appropriate model for forecasting and estimate model parameters, without the need for human intervention (Hyndman and Khandakar 2008). The most common automatic forecasting techniques used are the ARIMA and Exponential Smoothing (ETS) models. ETS models have been used extensively in forecasting demand for inventories (Billah, King et al. 2006). They have performed well in forecasting competitions against more complex models and are particularly good at forecasting seasonal data over short horizons (Hyndman, Koehler et al. 2002).

Automatic forecasting techniques can be too inflexible and do not allow for analysts with specialized knowledge to incorporate their expertise into the model effectively. This leads to difficulties in producing forecasts of high quality, thus the demand for high quality forecasts far exceeds the rate at which they are produced (Taylor and Letham 2017). In December 2016, Facebook released their forecasting model, Prophet. According to Taylor and Letham (2017), Prophet is a flexible model that can be adjusted by a large number of non-experts who have little knowledge about time series, however have expert knowledge on the data-generating process. Prophet allows for a large number of forecasts to be produced across a variety of problems and consists of a robust evaluation system that allows

for a large number of forecasts be evaluated and compared. This is Facebook's definition of forecasting at scale, and provides a solution to the problems that automatic forecasting pose (Taylor and Letham 2017).

The aim of this paper is to investigate the application of Prophet and ETS models in the prediction Bitcoin/ZAR. This paper will critically compare and evaluate ETS models and Prophet with the aim of determining if Bayesian forecasting methods are worthwhile.

2 A Brief Literature Review

2.1 Exponential State Space Models

2.1.1 Exponential Smoothing Methods

ETS methods originated in the 1950's and is the key work of Brown , Holt, and Winters (Gooijer and Hyndman 2006). It has been widely used for many applications in industry and has been the foundation of many successful forecasting methods (Gooijer and Hyndman 2006). Exponential smoothing techniques produce forecasts by calculating them as a weighted average of past observations. Observations that are most recent are more heavily weighted, with these weights decaying exponentially as the observations get older.

Exponential smoothing methods were originally classified by Pegel according to the components present in the time series (trend and seasonality), and whether they were additive or multiplicative in nature (Gooijer and Hyndman 2006). Gardner and Taylor further extended this classification to include a damped additive and multiplicative trend

respectively (Gooijer and Hyndman 2006). A taxonomy of exponential smoothing methods was later created by Hyndman et al. and extended by Taylor (Gooijer and Hyndman 2006). It considered all possible combinations of the trend and seasonality components thus producing 15 methods of exponential smoothing as seen in table 1.

Trend Component	Seasonality Component		
	None (N)	Additive (A)	Multiplicative (M)
None (N)	(N,N)	(N,A)	(N,M)
Additive (A)	(A,N)	(A,A)	(A,M)
Additive damped (A_d)	(A _d ,N)	(A _d ,A)	(A _d ,M)
Multiplicative (M)	(M,N)	(M,A)	(M,M)
Multiplicative damped (M_d)	(M _d ,N)	(M _d ,A)	(M _d ,M)

Table 1: Taxonomy of Exponential Smoothing Methods (Hyndman, Koehler et al. 2002)

The most common of these methods are Simple Exponential Smoothing (N,N), Holt's linear method (A,N) and Holt-Winters seasonal method (A,A).

Although exponential smoothing methods performed well and have been widely used in business, they did not have any statistical framework underpinning them (Gooijer and Hyndman 2006). They produced point estimates rather than prediction intervals and selection of a forecasting method was specific to the time series in question. Efforts were made by Box and Jenkins to create a statistical framework that underlies exponential smoothing methods and they found that forecasts produced by linear methods are special cases of ARIMA

models (Gooijer and Hyndman 2006). This however does not hold for the multiplicative methods.

2.1.2 State Space Models

The motivation behind state space models was the need for a statistical framework that underlies all methods of exponential smoothing (Gooijer and Hyndman 2006). Hyndman et al. (2002) provided this framework as an extension of Ord, Koehler, and Snyder's class of exponential state space models which underpinned some of the smoothing methods (Gooijer and Hyndman 2006). He proposed two state space models, one corresponding to an additive error and the other to a multiplicative error. By specifying a distribution for these error terms, he turned the previously deterministic methods into stochastic models. These two models were then applied to all 15 smoothing methods, producing 30 models in total.

The two state space models produce identical point estimates to each other as well as to the exponential smoothing methods, however they will generate differing prediction intervals (Hyndman and Athanasopoulos 2013). Each model consists of a measurement equation and a transition equation. The measurement equation shows the relationship between the observations and the unobserved components (level, trend seasonality), while the transition equation shows how the unobserved components change over time.

The introduction of the exponential state space model has allowed for model selection using information criteria rather than using an ad hoc method. It has also made the calculation of the likelihood easier and

simulation from the underlying model possible (Hyndman, Koehler et al. 2002).

2.1.3 Application and Performance

In a paper by Smith and Agrawal (n.d), Holt-Winters exponential smoothing (HWES) model was applied to time series data relating to data on patents. The patent data was classified into three different groups; monthly forecasts were made and the results were compared to forecasts produced by ARIMA models. To avoid over-forecasting, the trend of the HWES was damped to produce forecasts with a flat trend for later observations. The results showed that all models forecasted the data adequately. The HWES model fitted the data well and performed better than the ARIMA models in the case where the time series for the patent group was fairly stationary. Based on the forecasting results, it could not be concluded that a specific time series model is better for forecasting patent data.

Hassani et al. (2015) forecasted tourism demand in specific European countries over a short, medium and long term horizon. The ETS model was used to produce forecasts and was compared to six parametric and non-parametric techniques. The results showed that no single model is best for producing forecasts across all countries and time horizons. The ETS model together with the Neural Networks (NN) and Fractionalized ARIMA (ARFIMA) produced forecasts with the least amount of accuracy and are thus not relevant models for forecasting tourism demand in Europe. 24 steps-ahead error forecasts produced by the ETS model were found to be larger than the errors produced by the benchmark model - Recurrent Single Spectrum Analysis (SSA-R)

across all European countries.

2.1.4 Robustness

Gardner (2006) explained the robustness of ETS models by comparing them to equivalent models. Simple exponential smoothing which is equivalent to an ARIMA(0,1,1) model is the most robust method and has performed well in forecasting many types of time series. By comparing simple exponential smoothing to lower order ARIMA models, it is seen that errors that arise due to model specification is not as much of a problem as ARIMA models. Furthermore, ARIMA models tend to produce larger MSE due to model selection errors, and this is worsened if the errors are not normally distributed (Everette and Gardner 2006).

Satchell and Timmermann found that for time series with a finite history, the weights assigned to each observation in the simple exponential smoothing method are robust, provided that the variance of the random walk relative to the variance of the error component is not too small (Everette and Gardner 2006).

Forecasts produced by simple exponential smoothing have performed well in modelling annual sales of different products, as well as aggregated economic series with low sampling frequencies (Everette and Gardner 2006).

2.2 Prophet

2.2.1 Bayesian Generalized Additive Model

Prophet is the key work of Taylor and Letham (2017) and consists of a decomposable model with a component for growth, seasonality, and holidays. These components consist of linear and non-linear functions of time. This is similar to a Generalized Additive Model (GAM) which is a regression model that consists of non-linear and linear regression functions applied to predictor variables (Taylor and Letham 2017). Prophet, like GAM, frames the forecasting problem as a curve fitting exercise using backfitting to find the regression functions.

The growth component is modelled in a similar way to modelling population growths which uses a logistic growth model (Taylor and Letham 2017). Populations typically grow non-linearly (although the growth component could also be linear) up to an upper bound known as the carrying capacity, and remains constant thereafter. The rate at which the population grows changes over time, and this is accounted for by including changepoints in the model where the growth rate may be changed (Taylor and Letham 2017). These changepoints can be automatically selected and have a Laplace prior distribution placed on them (Taylor and Letham 2017). The parameter of the prior distribution can be used to adjust the growth rate of the changepoints. This allows non-experts with knowledge about events that may affect growth to use the parameter as a knob and adjust it to increase or decrease the number of changepoints (Taylor and Letham 2017). It also allows for the analyst to add changepoints which the automatic selection procedure may have missed (Taylor and Letham 2017). Furthermore, analysts may also specify the carrying capacity and adjust it based on their knowledge of the total market size (Taylor and Letham 2017).

The decomposable form of the model allows for components to be

easily added to it. This allows for multiple seasonality components with different periods to be added to the model. The seasonality component is modelled by a Fourier series, with the parameters of the Fourier series having a Normal prior distribution (Taylor and Letham 2017). The variance of the parameters can be adjusted by analysts to smooth the model and change how much of historical seasonality is projected to the future (Taylor and Letham 2017).

The name, date, and country of past and future holidays and events may be inputted by the analyst into a list (Taylor and Letham 2017). By specifying the country in which the events take place or the holidays occur, separate lists can be populated for global events/holidays and country-specific events/holidays. The union of the two lists are then used for forecasting. Like seasonality, a Normal prior distribution is placed on the parameters of the holidays, and the variance of the parameters can be adjusted by analysts to smooth the model (Taylor and Letham 2017).

2.2.2 Application and Performance

Taylor and Letham (2017) forecasted the number of events on Facebook using Prophet. The time series was impacted by holidays, had strong multi-period seasonality, and a piecewise trend. The forecasts produced by Prophet were compared to forecasts produced by common automatic forecasting techniques such as ETS, ARIMA, and the seasonal naïve model, as well as to simple models such as the naïve model. While the ETS and seasonal naïve model were quite robust, the ARIMA forecasts were fragile. No model besides Prophet accounted for the dips around holidays and the upward trend of the time series towards later observations. Hence, Prophet had lower forecasting errors

compared to the automatic forecasting methods.

2.2.3 Robustness

Prophet is a flexible model and has intuitive parameters that can be easily interpreted by human beings. The GAM is fitted quickly, allowing the analyst to interactively change the model parameters (Taylor and Letham 2017). Unlike ARIMA models, Prophet can produce forecasts over different scales and allows for missing values in the time series without the need for interpolation (Taylor and Letham 2017).

When a large number of forecasts are produced, manually identifying problematic forecasts becomes a time consuming and difficult task. Prophet provides a semi-automated forecast evaluation system that selects the best model that fits the data. When there are large forecast errors, the forecasts are flagged so that the analyst can explore the cause of the errors, identify and remove potential outliers and either adjust the model or choose a more appropriate model (Taylor and Letham 2017).

2.3 Evaluation of Forecasts

2.3.1 Common Methods

Model selection has historically been very subjective, with no techniques that offered mathematical rigour for choosing models (Smith and Agrawal). Progress has been made over time and many mathematical techniques have been introduced for model selection. Use of information criteria is one such method, with Akaike's information criterion (AIC) and Bayes information criterion (BIC) being the most

common (Smith and Agrawal). These information criteria minimize forecasting errors while penalizing it for overfitting. Information criteria should be used to compare models of similar structure only. To compare models of different structure, the most commonly used measures of forecast errors are the Root Mean Squared error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE). A lower value of these statistics indicates that the forecasts are better, with a MASE less than one indicating that the forecasts are better than those produced by a naïve model (Smith and Agrawal).

Smith & Agrawal (n.d) have used these methods in their paper when evaluating patent forecasts and making comparisons between the HWES and ARIMA models. In a paper by Hassani et al. (2015) they chose their ETS model based on the selection that R automatically made for the time series. They compared models of different structure using the RMSE and a modified Diebold-Mariano test. Hyndman et al. (2002) suggests that ETS models should be selected using the AIC rather than the MSE or MAPE. The AIC provides a way of choosing between models with an additive error and models with a multiplicative error since it is based on the likelihood function. The MSE and MAPE are not able to select between the two error types because both models produce the same point estimates. According to Gooijer and Hyndman (2006), the best approach of selecting models with different structure is the Diebold-Mariano test. The MSE is scale dependent and should not be used to make comparisons while the MAPE encounters difficulties when the time series values are close to or equal to zero.

2.3.2 Mincer-Zarnowitz Approach

Mincer and Zarnowitz (1969) proposed two methods of measuring forecast accuracy - an absolute and a relative measure. Absolute measures of accuracy measure the distance between actual and predicted values. If the mean distance is equal to zero, the forecast is said to be unbiased (Mincer and Zarnowitz 1969). To analyse absolute forecast errors, the observed values are regressed against the predicted values. A joint hypothesis test is performed to check if the intercept is equal to zero and the slope is equal to one. If the intercept is equal to zero, the forecasts are unbiased and do not overestimate or underestimate the data. If the slope is equal to one, the residual errors are uncorrelated with the predictions and the forecast is said to be efficient (Mincer and Zarnowitz 1969). The MSE is a measure that can be decomposed into the residual variance and two components, one reflecting the bias and the other reflecting the inefficiency of the forecast. If the forecasts are unbiased and efficient, the MSE will reduce to the residual variance (Mincer and Zarnowitz 1969). Absolute forecast measures cannot be used to make comparisons between forecasts with different scales or economic variables. Thiel states that the consequences of forecasting errors and how they impact decisions is of greater importance than the size of the forecasting errors (Mincer and Zarnowitz 1969).

Rather than using absolute measures of forecast accuracy, Mincer and Zarnowitz (1969) suggested the use of relative measures of forecast accuracy. Relative accuracy analysis allows for meaningful comparisons of different forecasting methods to be made. They proposed an index which considers the ratio of the MSE of the forecast to the MSE of a benchmark forecasting method. They used an

extrapolation of the data history as a benchmark, as it is a cost effective and accessible method, however the benchmark may be any method which is relevant for comparison. This ratio is known as the Relative Mean Square Error (RM).

The numerator represents a return which is inversely proportional to the MSE error of the forecasts, while the denominator represents the cost of forecasting which is inversely proportional to the MSE of the benchmark (Mincer and Zarnowitz 1969). Hence the ratio is representative of a rate of return index and ranks the performance of forecasts as such. If the forecasts are better than the benchmark, the RM will be less than one.

3 Further Reading

Some of the papers that will be read are:

Cho, V. (2003). "A comparison of three different approaches to tourist arrival forecasting." Tourism Management(24): 323–330.

MacDonell, A. (2014). "Popping the Bitcoin Bubble: An application of log--periodic power law modeling to digital currency."

Snyder, R. D., et al. (2002). "Forecasting for inventory control with exponential smoothing." International Journal of Forecasting(18): 5-18.

Taylor, J. W., et al. (2006). "A comparison of univariate methods for forecasting electricity demand up to a day ahead." International Journal of Forecasting(22): 1-16.

4 Methodology and Data

The daily closing prices of Bitcoin/ZAR will be obtained from Bloomberg, with the period spanning 2009 to 2017. The data will be cleaned and plotted in R Studio to look for trends and seasonality. The data will then be split into a training set and a validation set. This will allow for a portion of the actual observations to be used for comparisons against the forecasted values. It is important to choose the appropriate window size of the training set so that it fully represents the trend of the data and so that no model has an unfair advantage (Hassani, Silva et al. 2015). Hassani et al. (2015) chose the window size by looking for break points and setting the size of the window to the time at which the last break point occurred.

The data will initially be forecasted using the Simple Exponential Smoothing model and Prophet. The ETS models (and Prophet) will be evaluated using a Mincer-Zarnowitz regression test rather than the R packages which automatically determines the ETS model that fits the data. A Mincer-Zarnowitz approach will also be taken when comparing forecasts produced by Prophet and the ETS models.

5 Research Question

How does the latest Prophet forecasting model compare to the exponential state space model forecasts?

Sub-questions:

- When does Prophet perform well?
- When does the ETS model perform well?
- Could the Bitcoin bubble have been predicted by either forecasting method?
- Is there an opportunity in forecasting Bitcoin/ZAR?

6 Paper Structure

The proposed structure of the paper:

1. Abstract
2. Introduction
3. Background
4. Methodology and Data
5. Forecasting Methods
6. Results
7. Discussion
8. Conclusion
9. References
10. Appendices

7 Timeline

Dates	Tasks
Tuesday 9 th May	Submit final proposal Work on literature review
Monday 5 th June	Submit draft literature review Receive feedback on literature review
Monday 12 th June	Submit final literature review Start working on draft research project
Saturday 25 th June	----- Consolidation & Exams -----
Thursday 13 th July	Resume working on draft Learn all R coding required Understand the mathematics behind Prophet and ETS Complete draft
Monday 21 st August	First submission of draft Receive feedback on draft
Monday 11 th September	Final submission of draft Alter paper based on feed back
Monday 2 nd October	First submission of final paper Receive feedback
Monday 6 th November	Submit final paper ☺

References

- Billah, B., M. L. King, R. D. Snyder and A. B. Koehler (2006). "Exponential smoothing model selection for forecasting." International Journal of Forecasting(22): 239– 247.
- Everette, S. and J. Gardner (2006). "Exponential smoothing: The state of the art— Part II." International Journal of Forecasting(22): 637– 666.
- Gooijer, J. G. D. and R. J. Hyndman (2006). "25 years of time series forecasting." International Journal of Forecasting(22): 443– 473.
- Granvik, A. R. (2010). "Forecasting Exchange Rates."
- Hassani, H., E. S. Silva, N. Antonakakis, G. Filisk and R. Gupta (2015). "Forecasting Accuracy Evaluation of Tourist Arrivals: Evidence from Parametric and Non-Parametric Techniques."
- Hyndman, R. J. and G. Athanasopoulos (2013). "Forecasting: principles and practice."
- Hyndman, R. J. and Y. Khandakar (2008). "Automatic Time Series Forecasting: the forecast Package for R." Journal of Statistical Software **27**(3): 1-22.
- Hyndman, R. J., A. B. Koehler, R. D. Snyder and S. Grosea (2002). "A state space framework for automatic forecasting using exponential smoothing methods." International Journal of Forecasting(18): 439–454.

Mincer, J. A. and V. Zarnowitz (1969). "The Evaluation of Economic Forecasts."
3-46.

Smith, M. and R. Agrawal "A Comparison of Time Series Model Forecasting
Methods on Patent Groups."

Taylor, S. J. and B. Letham (2017). "Forecasting at Scale."