CHAPTER 2

LITERATURE SURVEY

Chapter 2

LITERATURE SURVEY

Many linear and non-linear models along with their variants are proposed in literature for forecasting time series data originating from various applications. The ARIMA and GARCH models are the popular linear models for time series forecasting. ARIMA is simple to apply and it assumes that the present data is linear function of past data points and past errors. It also assumes that errors are white in nature and requires that the data be made stationary before fitting a linear equation to the data.

2.1 ARIMA models

ARIMA models are popularized after Box and Jenkins (1970) who integrated the existing knowledge that time and devised developed a coherent, versatile three-stage iterative cycle for time series identification, estimation, and verification. This method is also known as the Box-Jenkins approach. Their contributions had an enormous impact on the theory and practice of modern time series analysis and forecasting [1]. In [11] quarterly automobile insurance paid claim costs are forecasted using econometric and ARIMA models. The forecast horizon is chosen as 8 quarters and the results indicate that forecast results are improved by adopting econometric models. In [12] the performance of

ARIMA model is compared with extended Wiener filtering for short term load forecasting in electric power systems. In [13] compares two alternative one-day-ahead forecasts of tomorrow's federal funds rate. The first forecast method employed is a simple random walk in which the forecast of tomorrow's federal funds rate is taken to be today's federal funds rate. The second alternative is forecasting using ARIMA model. These two forecasts are compared in terms of their general forecast accuracy and the decision support they provide to a financial institution and they found that forecast accuracy and the decision support superiority of the ARIMA forecasts is found to be quite small.

Box-Jenkins approach is employed to forecast US population totals in [14]. It is shown that the this approach is equivalent to a simple trend model when making long-range predictions and the prediction accuracy indicates that the Box-Jenkins method produces population forecasts that are at least as reliable as those done with more traditional demographic methods. ARIMA modeling with intervention analysis are used to predict incoming calls to telemarketing centers for the purposes of planning and budgeting in [15]. This prediction model is applied on both real data (Northern Italy) and simulated data. ARIMA models are used to forecast low volatile time series data such as climate change data, agricultural data etc. In [16] ARIMA model is used to forecast next-day electricity prices. In [17], ARIMA models are used to forecast the sugarcane production in Tamilnadu-India. To estimate Methane emission,

an ARIMA Model is used in [18]. ARIMA model is used to forecast employment information of Chinese computer industry in [19].

2.2 GARCH models

Engle in 1982 introduced ARCH model [20] which is known for forecasting highly volatile time series data. An improved version of ARCH model called generalized ARCH also known as GARCH model [21] was introduced by Bollerslev, Engle, and Nelson in 1994. Several extensions of GARCH model are found in literature. Some of the extensions are extensively surveyed in [22]. In [15] a simple (linear) GARCH(1,1) model is applied on Swiss market index and the results shows that the model provides a good parametrization for the daily returns. GARCH model is further explored and studied in [23] and showed that the presence of outliers has an impact on the forecast results. In [24] GARCH model is compared with other linear models and ANN in terms of onestepahead forecast accuracy. Forecast accuracy can be improved by augmenting models of volatility with measures of lagged volume. A variant of GARCH, Fractionally integrated GARCH (FIGARCH) is proposed in [25]. Further, in [26], the forecast content functions are compared from GARCH and FIGARCH models. The results shows that statistical forecasts tends to decline as the forecast horizon increases. They also shows that forecasts by autoregressive projection on past realized volatilities seems to provide improved results in terms of forecast accuracy when compared to that of GARCH, estimated by quasi-maximum

likelihood, and FIGARCH models.

In [27] GARCH models are used to forecast hourly prices in the deregulated electricity markets of Spain and California and the results outperforms ARIMA model when volatility and price spikes are present.The ARIMA model is used in forecasting solar radiation at high resolutions in [28]. ARIMA is applied on various time series data like electricity price [29] [16], sugar price [17], stock market data [30], [31], wind speed [32], monthly water inflow [33], wireless network traffic data [34] etc, and is good in understanding the dynamics of the data.

ARIMA and GARCH models are used to forecast wind speed in [35]. This research uses five different GARCH approaches and each consists of an original form and a modified form, GARCH-in-mean (GARCH-M). Ten different model structures are evaluated, based on the 7-year hourly wind speed data. The highlights the efficiency of this model approaches in catching the trend change of the mean and volatility of wind speed. Seasonal GARCH models are used on internet traffic data for prediction as shown in [36]. Auto Regressive Conditional Heteroscedastic (ARCH) models can also be used in prediction. The use of ARCH and GARCH models in econometrics is highlighted in [37]. Volatility forecasting using GARCH on Hong Kong stock exchange is presented in [38], which emphasizes that the GARCH models with non-normal distribution shows better volatility forecasting performance in comparison with historical models. In [39] [40], AR and GARCH models are used for forecasting financial data.

A prediction technique can be employed on the raw data, or the data can be pre-processed before the technique is employed. Such a preprocessing can improve the prediction accuracy. For example, in [8], to predict global temperature values, trend based ARIMA and wavelet ARIMA are used. It is shown that trend based ARIMA outperforms simple ARIMA and wavelet based ARIMA outperforms trend based ARIMA. Similarly in [9], it is shown that wavelet based ARIMA outperforms simple ARIMA model in case of forecasting day ahead electricity prices. Another study also highlights the use of wavelet transform to improve the prediction accuracy of a time series data. When time series data is decomposed into a time dependent sum of frequency components, it is easy to capture seasonalities with time-varying period and intensity. This type of forecasting eventually improves the forecasting accuracy as shown in [41]. In [42] wavelet decomposition method is used for time series forecasting in business-field. using ARMA model near-periodicity, nonstationarity and nonlinearity is captured from business short-term time series. The wavelet decomposition is applied to BP networks and an ARMA model. With this approach the forecast results obtained are more accurate than the conventional techniques, like those only using BP networks or ARMA models. In [9] wavelet transform and ARIMA models are used to forecast day-ahead electrify prices of Spain.

Many Hybrid models are derived in literature for time series forecasting. The hybrid models consider the strengths of individual models and combine them in some form to get better prediction accuracy. In

[43] ARIMA and GARCH models are used to model rainfall Time Series. The results shows that the composite ARIMA-GARCH model captures the dynamics of the daily rainfall series accurately than individual models, and the seasonal ARIMA model is suitable for forecasting monthly average rainfall data considered for testing. ARIMA is used with GARCH model for forecasting daily load and maximum electricity demand estimation as seen in [2], which provides a better long term forecasting. In [44], traffic modeling and prediction using ARIMA and GARCH models are discussed.Some times GARCH models are inappropriate to model as they assume Gaussian distribution. In such cases student-T distribution can be assumed. This fact is further investigated in [3]. Spectral analysis is used to forecast real time traffic flow in [45], which includes the modal functions associated with a covariance matrix constructed from historical flow data for forecasting traffic flow. An investigation of one-step-ahead forecasting performance of GARCH models for the $S\&P-100$ stock index is presented in [46]. The period considered for investigation from 1 October 2001-15 September 2003. Performance of different GARCH variants are studied on the data.

A prediction technique based on the Moving Average (MA), Exponential Smoothing (ES), ARIMA, and ANN model is proposed in [47]. Models based on ANN [48],[49],[50], [51], [52] are also applied for prediction of time series data. In [53] the relationship between ANN performance and ARIMA-GARCH model parameters are presented. This model uses a feed

forward ANN which is trained with the Levenberg-Marquardt (LM) algorithm, through Monte Carlo simulations. Prediction models can also be embedded in the form of an architecture. Architecture can contain many layers which represent various processing steps used on the data for prediction in a sequence. A combination of ARIMA-GARCH model is proposed for Short-time traffic flow prediction [54]. The proposed model combines linear ARIMA model with nonlinear GARCH model to capture both conditional mean and the conditional heteroscedasticity of traffic flow series. It is observed from the results that the hybrid model proves more prediction accuracy when compared to that of standard ARIMA model.

2.3 ANN models

Neural Networks have gained more popularity for past several years. The research on ANN earlier quoted earlier in [55]. One of the major application of neural networks is forecasting as mentioned in [49]. ANN can be useful for nonlinear processes, but they are difficult to fit since it has unknown functional relationship as mentioned in [56]. This research compares ARMA and ANN forecasting performance and presents the conditions under which neural networks could be superior to linear models. Literature has numerous examples and documentation for the success of ANN in forecasting However, two editorials [57] [58] question the importance given by the researchers for ANN in solving various

statistical problems especially the forecasting time series data. The research work [59] suggests that ANN is a suitable model for forecasting high frequency data. An attempt to provide a balanced assessment on ANN, this research provides a review on the literature comparing ANN and statistical models such as regression-based time series forecasting. The empirical studies studies find ANN comparable with their statistical counterparts. Modeling and forecasting trend time series is presented in [60]. A detailed discussion on modeling trend time series using neural networks is presented. Different trend patters like linear, nonlinear, deterministic, stochastic, and breaking trend are considered for modeling.

ANN is advantageous compared to ARIMA in many applications because ANN does not assume linearity. It is capable of fitting a nonlinear function to the given data and does not need the data to be made stationary. It is also adaptive in nature. Due to these reasons ANNs have become more popular. The research work of [61] stresses upon the efficiency ANNs over ARIMA in forecasting TSD. A three-layer back propagation (BP) neural network is used for electricity price forecasting in [62]. The results shows that the electricity price in the deregulated markets is dependent strongly on the trend in load demand and clearing price. Some of the contributions of this research include, firstly accuracy of ANN and data mining approach is reduced when more factor analysis is incorporated, secondly, higher accuracy in the model result is observed on approaching closer to the actual occurrence of financial distress. The third contribution is factor analysis increases the error of

classifying companies that are in a financial crisis as normal companies. In [63] Neural networks are used to predict earthquakes in Chile. ANN was able to model a wide range of TSD compared to ARIMA, as they are capable of modeling non linear variations. Compared to ARIMA, these TSD were accurately predicted with ANN.

ANNs have been applied to electricity demand data [62], financial data [50] river flow data [64], and network data [65], for prediction. In [66] ANNs are used to forecast the maximum daily precipitation for the next coming year in Athens, Greece. Satisfactory results are obtained when this model is used for forecasting one year ahead daily precipitation totals. A variant of ANN called spiking neural network is used to forecast [67] short-term electrical load of Victoria State, which shows better prediction accuracy than SVM hybrid models. In [52], it is shown that ANN outperforms the transactional AR models in TSD prediction. A three-layered feed forward ANN with Levenberg-Marquardt algorithm is used to predict next-week electricity prices of mainland Spain and California electricity market in [68]. The Zhang's model [69] highlights the use of ANN approach for forecasting quarterly TSD with a large data set of 756 quarterly TSD. In [70], ANNs are used for forecasting seasonal TSD.

Zhang's idea [5] of decomposing the TSD in to linear and non-linear components is employed in several research works. The purpose such decomposition is to split the TSD in to linear and non-linear TSD series, later different prediction models can be used to model each one

of them. In [71] after decomposing the TSD, the linear component is modeled using ARIMA and the non-linear component is modeled using support vector machines. Experiments are conducted on several stock market TSD and the results reveal that the proposed hybrid model has better prediction accuracy compared to that of individual models that is ARIMA and Support Vector Machines(SVM). In [72] TSD is decomposed in to two components, the linear component is modeled using generalized linear auto regression (GLAR) and non-linear component is modeled by ANNs. This hybrid model is applied to forecast foreign exchange rates. The results reveal that proposed hybrid model is better than the individual models considered for comparison. A hybrid ARIMA-ANN model for forecasting TSD is presented in [73] which uses fuzzy logic-based system rules in order to obtain the order of differencing for ARIMA model. The prediction accuracy over the individual models is highlighted using the simulated data sets. Combination of principal component analysis (PCA) and ANN are used in forecasting ground-level ozone in [74]. A hybrid model designed using ARMA and ANN [75] adopts similar decomposition of TSD. In [76] day-ahead daily flows of Portuguese watersheds is forecasted using the computational neural networks (CNNs) combining with ARIMA model in the similar manner. Solar radiation TSD is forecasted using a hybrid ARIMA and Time Delay Neural Network (TDNN) in [77]. In [78] a hybrid model is devised by combing Recurrent Neural Networks (RNN) and ARIMA models. In the hybrid model proposed in [79], ARIMA is fit on the given TSD, then dividing given TSD by the

ARIMA predicted TSD, the non-linear component is obtained, which is predicted using ANN. This model showed almost similar prediction accuracy as that of Zhang's hybrid [5]. In [79], a multiplicative model was proposed (Li Wang et.al), which is in contrast to the additive model of Zhang. The results showed that it is no less in comparison to the application of additive Zhang model.

In [80] ARIMA models are used in devising a new hybrid method for improving the performance of ANNs. In this approach the which is based on the Box-Jenkins methodology, TSD considered as a nonlinear function of several past observations and random errors. In order to generate the necessary data ARIMA is used and to find out the model order ANNs are used. A comparison of forecasting performance of ANN over ARIMA and hybrid models are presented in forecasting wheat price in Chinese market in [81]. The results shows that ANNs can predict better compared to other models considered where the bench mark set is ARIMA model. Hybrid models using ANN and AR models are deployed in [82] to forecast the monthly streamflow data from the Colorado River in USA. The results reveal that the ANNs can capture the hidden relationships in the considered data better compared to that of conventional AR models.

In [83] a multi-layer perceptron model is devised for predicting hourly concentrations of nitrogen dioxide in Helsinki. A parallel genetic algorithm (GA) is used for designing the high-level architecture and for selecting the inputs which improves the prediction model performance

better. However, this process seem to be expensive which is a limitation of this model.

A flood forecasting model is built using ANN for Dadu River data in China in [84]. Since a technique with self-learning, self-organizing with fault tolerance abilities is very much important for flood forecast. This model offers better prediction accuracy when compared to other conventional prediction models. The experimental results have reported that the maximum error of only 0.02. The Zhang's model [69] highlights the use of ANN approach for forecasting quarterly TSD with a large data set of 756 quarterly time series. Data has 48 systematic models with a variety of input variables. Some of the findings of this research work are as follows. From the perspective of variable selection, different ANN models perform differently and the ANN performance depends on data preparation. Performance was high when the model is simple that is, when both number of input nodes used and the number of hidden nodes selected are simple. By including seasonality information the performance of ANNs does not improve. Generally, ANN modeling is not helpful in improving forecasting performance in such case.

A forecasting model with algebraic forecasting which uses internal smoothing is proposed in [85]. This technique is based on skeleton algebraic sequences identification and also by finding the near-optimal balance between algebraic variability and smoothness of moving averages. This method is efficient even when TSD is short but sufficient enough to train the model. The data from Spanish electricity market is

considered for evaluating the model.

If there is an interval in the data then the forecasting becomes more difficult since the data is missing at random intervals. An attempt to forecast such data is made in [86]. This research discusses four different approachers-Autoregressive (AR), ARIMA, ANN, and hybrid ARIMA-ANNs. In [31] a text mining approach with the combination of ARIMA and support vector machines are used to forecast stock market data. Six different company data is used to evaluate the proposed method. The results show that this approach improves the prediction accuracy. In [83] a multi-layer perceptron model is devised for predicting hourly concentrations of nitrogen dioxide in Helsinki. A flood forecasting model is built using ANN for Dadu River data in China in [84].

A hybrid neural network model is proposed in [87] to forecast monthly streamflow data at Colorado River at Lees Ferry, USA. Preprocessing techniques such as de-trending and de-seasonalisation, are carried out before passing the time series data to the ANN. The results shows that by combining the strengths of the conventional and ANN techniques provides a robust model which is capable of capturing the non-linear nature of the complex time series and thus producing improved prediction results. In [88], new classification and feature extraction techniques were proposed for electrocardiography data. These preprocessing steps can be applied to the raw data to obtain more accurate predictions. In this research, the basic ARIMA model was chosen as the linear prediction model. In [89] to forecast end-use demand three conventional ANNs are

used: two feed-forward back propagation networks and one radial basis function network. The accurate prediction are obtained using a sigmoid activation hidden layer and linear activation output layer.

Other than ARIMA and ANN, many other prediction models are available in literature. Some of them are SVM based [90], fuzzy based [91][92], Hidden Markov Model (HMM) [93] [94], etc. Modified forms of ARIMA like SARIMA [95], FARIMA [96] are also available. Spectral techniques based on SVD are proposed in [97] and the references there-in. A statistical wind forecasting tool using PCA is proposed in [98], which is applied on wind time series data. In [99], a layered architecture is proposed for financial time series data. This architecture is shown to have been applied on various financial time series and the success of the architecture is discussed with regard to financial data.

There is no universal model which can suit in all applications. The prediction accuracy can be improved if two different models are applied to the same data rather than a single model. Most of the times it is seen that instead of one model, if hybrid models are used prediction accuracy improves, but if the hybrid uses many decompositions and many models, then accuracy will degrade after a limit and the model will no longer be successful. So hybrid models must not include too many models. They should contain limited number of individual models to retain the model simplicity and prediction accuracy. The models proposed in the literature tend to combine the advantages of two or more models to design a best model.