PREDICTIVE MODELS FOR AIR SHOW TICKET SALES

A Thesis

Presented to the Faculty of the Department of Mathematical Sciences

Middle Tennessee State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Mathematical Sciences

by

Ye Fang

May 2018

Thesis Committee:

Don Hong, Chair

Dennis P. Walsh

Qiang Wu

ACKNOWLEDGMENTS

I would like to acknowledge a few people for helping me with this thesis. I would like to thank Dr. Don Hong, who has been very helpful and patient through the entire process. I am also very grateful for Dr. Qiang Wu for many helpful discussions and comments. To Dr. Dennis P. Walsh, thank you for your patience and valuable suggestions. I appreciate the Middle Tennessee State University Actuarial Science Program for providing me the opportunity to work on the air show tickets case study started in the spring of 2017 with Mr. Stephen Henderson, an actuarial student in the MSPS program. I am grateful to Stephen Henderson for his contribution to the three-segment latent class Weibull modeling. The air show tickets case study has inspired my interest in predictive analysis.

ABSTRACT

Promoters and companies often want to forecast the success of their event while their tickets are still on sale. Prediction and good estimation of ticket sales will allow companies to see if they need to plan for things in advance, such as advertisements for their event and the arrangements of services for the event. The purpose of this study is to predict the ticket sales for air shows put on by a ticket-selling company. This study uses the ticket sales data of four past events of the Great Tennessee Air Show. We study two kinds of prediction models, the three-segment latent class Weibull model and dynamic neural network model. We then compare the prediction results of both models. We conclude that the three-segment latent class Weibull model is more suitable for long-term prediction, and the dynamic neural network model is better for short-term forecasting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Ticket-selling companies want to know how their events are doing in sales. They especially want to have an estimation of how many tickets will be sold before the event happens so that they can determine if they need to allocate resources towards advertising and arranging of other services. Some of these services include food purchases, contact of law enforcement for traffic, and staffing for the show. Jacob Suher [1] provided some reasons into why predicting ticket sales is a critical element of event planning and discussed how to determine the specific modeling needs. Air shows are usually two-day events for air performers. Since the air performers typically put on the same act for both days, consumers can choose either day to watch the air show.

The purpose of this air show ticket study is to predict the ticket sales for air shows for a company by using their previous air show ticket data. In this thesis, we try to adapt two different models to forecast tickets sales. The first is the three-segment latent class Weibull model. The Weibull distribution [2] is named after Swedish mathematician Waloddi Weibull and is used in survival analysis, wind speed [3], weather forecasting and so on. The latent class (LC) [4] modeling was initially introduced by Lazarsfeld and Henry as a way of formulating latent attitudinal variables. We incorporate Weibull distribution into latent model for air show ticket sales prediction. Our second model is the dynamical nonlinear autoregressive (NAR) neural network model. The NAR neural networks [5] are powerful computational models for modeling and forecasting nonlinear time series. We find the three-segment latent class Weibull model gives a better prediction for the upcoming event before ticket sales begin. However, the dynamic NAR neural network model is better suited for predicting ticket sales after sales have began.

We organize this thesis as follows. The analysis of raw data is given in Chapter 2. The three-segment latent class Weibull model and experimental results are shown in Chapter 3. The algorithm and modeling of dynamical NAR neural network and results are presented in Chapter 4. Some comparative analysis and conclusions are summarized in Chapter 5.

CHAPTER 2

DATA ANALYSIS

The data for the air show study consists of the Great Tennessee Air Shows in 2011, 2012, 2014, and 2016. A few lines of the data from 2011 are shown in the table below. This data has eight variables: event ID number, order ID number, event name, ticket ID number, ticket name, price, the time when the ticket was purchased, and the time when the event started. The various ID numbers, event names, and prices are not used in the analysis.

Order ID	Event ID	Event Name	Ticket ID	Ticket Name	Price	Purchased UTC	Event Start UTC
1010006	755	The Great...	1305	Children 4-12 One...	8	4/17/2011 06:31 AM	5/7/2011 01:00 PM
1010013	755	The Great...	1305	Children 4-12 One...	8	4/17/2011 06:54 AM	5/7/2011 01:00 PM
1010014	755	The Great...	1304	Adult One Day...	15.99	4/17/2011 03:49 PM	5/7/2011 01:00 PM
1010014	755	The Great...	1304	Adult One Day...	15.99	4/17/2011 03:49 PM	5/7/2011 01:00 PM
1010014	755	The Great...	1305	Children 4-12 One...	8	4/17/2011 03:49 PM	5/7/2011 01:00 PM
1010015	755	The Great...	1317	Flight Line...	95.99	4/17/2011 05:07 PM	5/7/2011 01:00 PM
1010015	755	The Great...	1317	Flight Line...	95.99	4/17/2011 05:07 PM	5/7/2011 01:00 PM
1010016	755	The Great...	1304	Adult One Day...	15.99	4/17/2011 05:31 PM	5/7/2011 01:00 PM
1010016	755	The Great...	1305	Children 4-12 One...	8	4/17/2011 05:31 PM	5/7/2011 01:00 PM
1010017	755	The Great...	1316	Premium Box Seats	25.99	4/17/2011 06:19 PM	5/7/2011 01:00 PM

The sequence of daily or weekly ticket sales is a time series. We focus on the number of tickets sold at a specific time, therefore, another time parameter is added which is the time interval between the start of the event and the time a ticket was purchased. This parameter is measured in days. If there is a difference of zero, then the person bought their ticket right at the beginning of the event. The company usually opens up sales for these tickets about 180 days before the event. However, each of the events varies in their maximum time differences, some above and below 180 days. Each of the differences were shifted until the maximum time difference was set to 180. This allowed for each of the events to be standardized up to 180 days.

In order to avoid any negative differences, the time when the first ticket was sold became the event's start time. The numbers in the new time variable were also reversed to create another time variable in order to show the day of a ticket sale. Therefore, a difference of 180 would become the first day of sale in the new dataset. In this case study, the days of sales were converted into weeks of sales. All of the partial weeks were rounded up so that the sale would range from 1 to 26 weeks of sales. Each ticket that had the same week as another ticket sold was added together. The final dataset was completed and ready to be used for plotting. A few lines of the correspondingly dataset from 2012 are shown in the table below.

Year	Week of Sale	Count
2012	1	68
2012	2	77
2012	3	92
2012	4	97
2012	5	108
2012	6	110
2012	7	115
2012	8	124
2012	9	142
2012	10	164
2012	11	172
2012	12	174
2012	13	204
2012	14	227
2012	15	287
2012	16	337
2012	17	389
2012	18	425
2012	19	493
2012	20	572
2012	21	697
2012	22	877
2012	23	1184
2012	24	1814
2012	25	4116
2012	26	10359

The graph of the actual cumulative sales of 2011,2012 and 2014 are shown below.

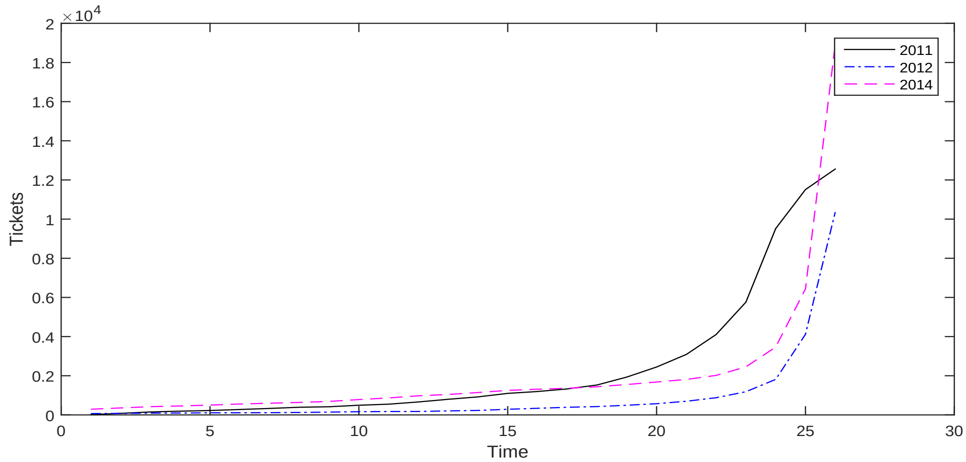

Figure 1: Actual Sales

CHAPTER 3

THE THREE-SEGMENT LATENT CLASS WEIBULL MODEL

In this chapter, we consider a three-segment latent class Weibull model to fit and to predict for air show ticket selling. The latent class model describes the air show ticket selling by separate segments, and then improves model fit relative to the single segment model.

3.1 The Three-segment Latent Class Weibull Model

A statistical distribution is used to fit curve [6]. The best way to determine the best model for this dataset uses the hazard rate function. The hazard rate function is given by:

$$h(t) = \frac{f(t)}{1 - F(t)}, \tag{1}$$

where $F(T)$ is the cumulative distribution function (CDF), $f(t)$ is the probability density function (pdf) of the corresponding random variable.

Since the purpose of the hazard rate is to determine the purchase rate in regards to time, it can be used to determine the best statistical distribution for this model. The Figure 1 shows that the curve is monotonically increasing as the event is approaching.

Even though there are many distributions that have similar hazard rates, the best distribution for this model is the Weibull distribution. Its CDF and hazard rate function are as shown:

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}} \tag{2}$$

$$h(t) = \alpha \lambda^{\alpha} t^{\alpha - 1}, \tag{3}$$

where α is the shape parameter and λ is the scale parameter. If the shape parameter α is greater than 1, then the hazard rate function is monotonically increasing.

However, only two parameters would not be sufficient for the Weibull distribution to fit the curve. We considered the latent class model. The latent class model allows heterogeneity in the shape parameter α and the scale parameter λ. Thus, a three-segment latent class Weibull model [4] is used for this research. Functionally, the same Weibull distribution models are used, but now with three discrete segments of parameters. The entire curve is made up of a percentage of each segment. The CDF is shown below:

$$F(t) = p_1(1 - e^{-(\lambda_1 t)^{\alpha_1}}) + p_2(1 - e^{-(\lambda_2 t)^{\alpha_2}}) + (1 - p_1 - p_2)(1 - e^{-(\lambda_3 t)^{\alpha_3}}), \tag{4}$$

with the following parameters:

α_1 is the shape parameter of the first segment,

λ_1 is the scale parameter of the first segment,

p_1 is the proportion of tickets sold of the first segment,

α_2 is the shape parameter of the second segment,

λ_2 is the scale parameter of the second segment,

p_2 is the proportion of tickets sold of the second segment,

α_3 is the shape parameter of the third segment,

λ_3 is the scale parameter of the third segment.

Since this equation would only reach up to 1, the expected frequencies N of this model would need to be calculated. It represents the max predicted number of tickets sold during the time period. The final formula of three-segment latent class Weibull distribution used in this study is shown as:

$$\hat{N}(t) = N[(p_1(1 - e^{-(\lambda_1 t)^{\alpha_1}}) + p_2(1 - e^{-(\lambda_2 t)^{\alpha_2}}) + (1 - p_1 - p_2)(1 - e^{-(\lambda_3 t)^{\alpha_3}}))]. \tag{5}$$

The parameter estimates can be calculated since the data has been condensed and the distribution has been found. The best technique was to find the Maximum Likelihood Estimate (MLE) for each parameter. This can be achieved by finding the log-likelihood function for the data. This data was treated as complete, grouped data. The likelihood function for this type of data is as shown:

$$L(\theta) = \prod_{j=1}^{K} [F(t_j|\theta) - F(t_{j-1}|\theta)]^{n_j}, \tag{6}$$

where K represents the total number of data, in this study $K = 26$; t_j represents each week of the ticket sales and n_j represents the number of tickets sold in that week, for $j = 1, 2, \cdots, K$; θ represents the estimated parameter, $\theta = \alpha_i, \lambda_i$ and p_i, for $i = 1, 2, 3$.

The logarithm of the likelihood function is:

$$l(\theta) = \sum_{j=1}^{K} n_j \ln[F(t_j|\theta) - F(t_{j-1}|\theta)]. \tag{7}$$

The log-likelihood function must be maximized in order to find the MLE. Since it is unrealistic to find the MLEs for an 8-parameter model by hand, they were found by using MATLAB. This study used MATLAB in order to maximize the log-likelihood function.

There are certain interpretations for each of the parameters. The shape parameter (α) has three conditions that can affect the curve. If the shape parameter (α) is less than 1, then the purchase rate decreases and results in an early life. If the shape parameter (α) is close to or equal to 1, then the purchase rate is constant, resulting in an approximate of an exponential distribution. If the shape parameter (α) is greater than 1, then the purchase rate increases resulting in a wear-out life. As the scale parameter (λ) increases and the shape parameter (α) remains constant, the maximum of the probability density function increases. If any of the proportion parameters (p) increase, then the model becomes more like a single Weibull distribution with the respective parameters of that segment. The ticket quantity (N) is the amount of tickets that are projected to be sold for an event.

3.2 Experimental Results

Since there are only 3 groups of train datasets, to even out fluctuation, we added a new dataset which is the average of 2011, 2012 and 2014, giving us 4 groups of samples to be fitted. The graph of the actual sales of 2011, 2012, 2014 and the average are shown below.

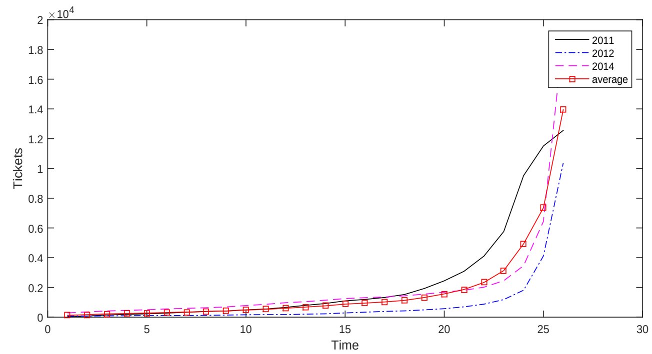

Figure 2: Actual Sales with The Average Dataset

The following are graphs of all train datasets fitted with the latent class Weibull Model by MATLAB:

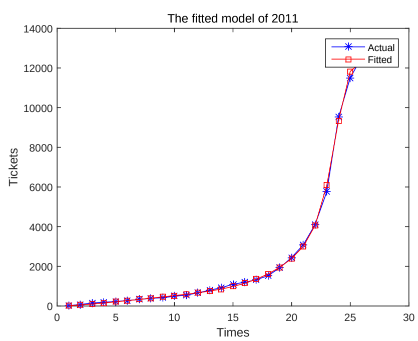

Figure 3: The Fitted Curve of 2011

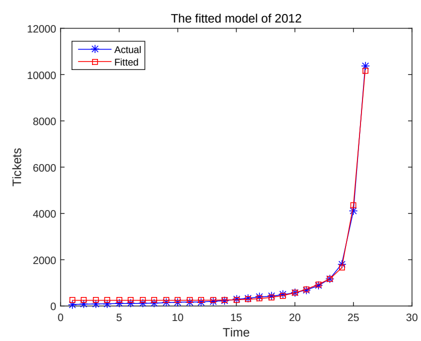

Figure 4: The Fitted Curve of 2012

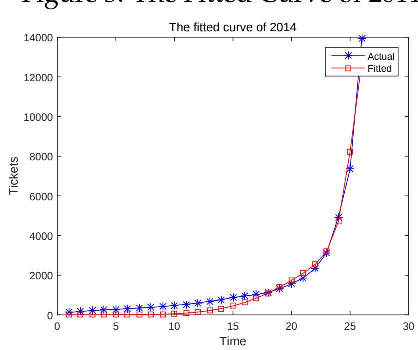

Figure 5: The Fitted Curve of 2014

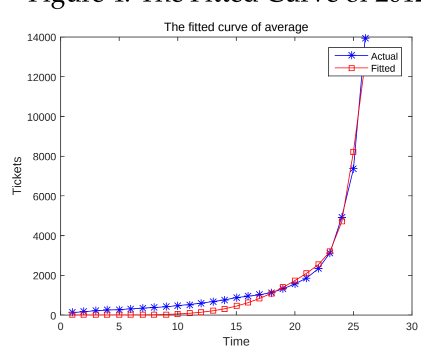

Figure 6: The Fitted Curve of The Average

Using the three-segment latent class Weibull distribution model prediction, we find the parameters of the three-segment latent class Weibull distribution, obtain values, and fully describe the fitted curve. Therefore, the main task is to find the parameter estimates for 2016 under the three-segment latent class Weibull distribution model. In this study, we assume that there is no other influence except the time effect. The method to calculate the parameters for 2016 uses the average of the 4 groups of fitted parameters, estimating each parameter one at a time. The results of all the parameters are shown in the following:

Table 1: Fitted Parameters of Train Datasets

	p_1	λ_1	α_1	p_2	λ_2	α_2	λ_3	α_3	N
2011	0.3021	0.0449	6.8377	0.0769	0.0834	1.6139	0.0419	26.0591	12,557
2012	0.1470	0.0406	6.2746	0.0220	0.0614	0.6754	0.0390	46.1624	10,359
2014	0.3175	0.0418	7	0.0985	0.5	2	0.0397	35	18,925
3 yar-average	0.2616	0.0461	5.5874	0.0016	0.0684	1.9560	0.0395	32.3640	13,747

Table 2: Estimated Parameters of 2016

	Segment 1	Segment 2	Segment 3
α	7.3715	1.3835	38.8975
λ	0.0431	0.1908	0.0401
p	0.2492	0.0648	0.686
N	13,947		

The predictive model of air show ticket sales for 2016 is shown in the figure below:

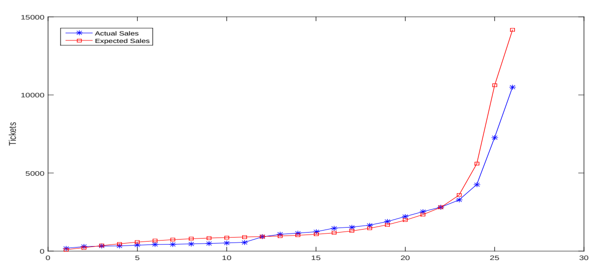

Figure 7: Predictive Model of Ticket Sale in 2016

3.3 Summary

We made use of the three-segment latent class Weibull distribution model for air show ticket sales prediction. In our predictive model, all the shape parameters (α) are greater than one. Although the scale parameter differs, with the second parameter being slightly greater than the other two parameters, this difference helps to fit the flat part of the curve. The curve for the later time periods when sales are greatest is based on the third proportion parameter.

The prediction model is based on the hypothesis that there is no other influence except time effect. Although other factors can affect ticket sales. Our predictive curve for 2016, is performed well using only time as a factor.

CHAPTER 4

DYNAMIC PREDICTION BASED ON NAR NEURAL NETWORK

4.1 NAR Structure

Artificial neural networks simulate biological neural networks and have many uses, including non-linear function approximation, time series prediction, and system control. There are two kinds of dynamic neural networks, the nonlinear autoregressive exogenous model(NARX)[7] is used for input-output modeling of nonlinear dynamical system; the nonlinear autoregressive models (NAR) predicts future values based only on several past values. The NAR neural network has great fitting ability for a nonlinear time series. NAR networks have three layers: the input layer (consisting of signal source nodes), a hidden layer (with a input delay function), and a linear output layer. The input layer is connected to the hidden layer by non-linear transformation. The hidden layer completes the nonlinear transformation of the input vector. On the other hand the transformation from the hidden layer to the output layer is linear, that is, the output of the network is the linear weighted sum of the outputs of the hidden layer nodes [8]. The basic structure is shown in the figure below.

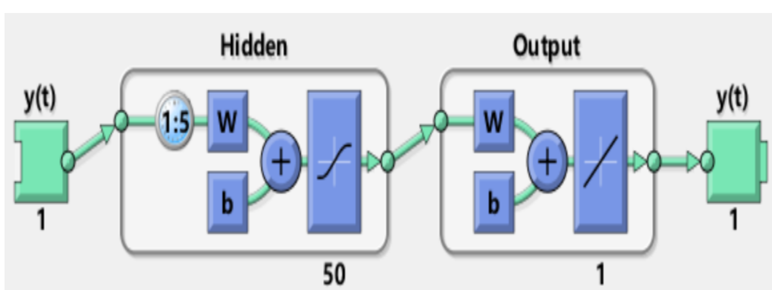

Figure 8: NAR Schematic Structure of The Neural Network

In the figure, W denotes the weighting coefficient in the network, b denote bias term, and 5 denotes the delay order.

The model of the NAR neural network is expressed as:

$$y(t) = f(y(t-1), y(t-2), \cdots, y(t-d)), \tag{8}$$

where d is delay order.

From the equation in (8), the value of $y(t)$ is determined by the value of $y(t-1)$, $y(t-2)$, $\cdots$, $y(t-d)$, that is, using the past known values to infer the current value which reflects the continuity of the process. If there exists a time series $y = \{y_i \mid y_i \in \mathbb{R}, i = 1, 2, ..., L\}$, we use the first d values to predict the next m values. The data can be divided into k segments of length $d + m$ with certain overlap, and each segment can be regarded as a sample. We obtain k samples, where $k = L - (d + m) + 1$, and we use these k samples to create the NAR neural network.

4.2 Establishment of NAR Neural Network

First, to build the model, we set the d delay order parameter and the number of neurons. Second, we train the NAR neural network in the openloop model using the Levenberg-Marquardt training function, and stop training when the network meets certain requirements. Next, we change the openloop model into the closeloop model, and then we obtain the predicted value. Additionally, we use squared error (R) and mean square error (MSE) to evaluate the network performance, we also use the error autocorrelation curve to judge whether the network is feasible.

Model steps:

- Set the delay order d, and the number of hidden neurons M.
- Set 70% of data for training sets, 15% for validation sets, and 15% for testing sets. Use the neural network toolbox in MATLAB to train the network.
- Make use of this network to predict air show ticket selling.

Table 3: Predictive Index of 2011 of NAR Network

Index	Training	Validation	Testing
MSE	350.8617	21208.8447	3163.3202
R	1	0.9984	0.9658

In this study, since we have 26 weeks of each year, and there is a minimal change in the first 9 weeks, we use $d = 9$ for the dynamic NAR neural network model.

The results of data in 2011 are shown in the following graphs:

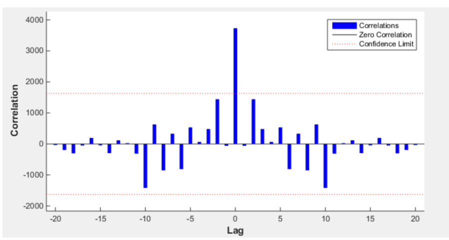

Figure 9: Autocorrelation of Error of 2011

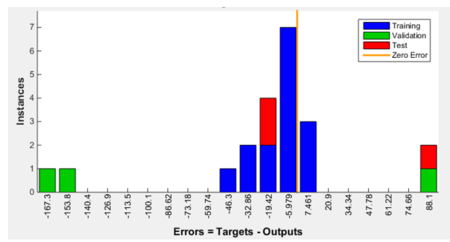

Figure 10: Error Histogram With 20 Bins of NAR Network of 2011

4.3 Prediction Modeling

The prediction method of the NAR model is one kind of a recursive prediction method [9], and the main idea is the cyclic use of the 1-step forward prediction value [10]. The basic step is :

$$y(t_n) = f(y(t_n - 1), y(t_n - 2), \cdots, y(t_n - d)). \quad (9)$$

If we have obtained the first predictive value $y(t_n)$, to predict $y(t_n + 1)$, we substitute $y(t_n - d)$ by $y(t_n)$. We obtain a new sample $y(t_n), y(t_n - 1), y(t_n - 2), \cdots, y(t_n - d + 1)$. We repeat the basic step based on the new sample, and then, we obtain the second predictive value $y(t_n + 1)$. We repeat this loop until we obtain all the predictive values we need. In this study, since $d = 9$, the first predictive value is $y(10)$ and we need to predict the next 17 values. Therefore, the whole recursive process is represented by equation:

$$\begin{cases} y(10) = f(y(9), y(8), \cdots, y(1))) \\ y(11) = f(y(10), y(9), \cdots, y(2)) \\ \vdots \\ y(26) = f(y(25), y(24), \cdots, y(17))). \end{cases} \tag{10}$$

We use the MATLAB function $removedelay()$ to predict the 1-step forward prediction value. This function returns the network with input delay connections decreased, and output feedback delays increased, by the specified number of delays n. Here, let $n = 1$. Then, using loop statements in MATLAB, we can implement the above algorithm.

There are various kinds of factors that will influence air show tickets selling, for example market environment, weather and so on. However, we have reason to believe, the closer the time, the greater the influence. Therefore, we add a larger weight for the closer predictive value. The final predictive value at time t is the weighted average :

$$y(\hat{t}) = \frac{w_1 y_1(t) + w_2 y_2(t) + w_3 y_3(t)}{w_1 + w_2 + w_3}, \tag{11}$$

where $w_1 = 1, w_2 = 2$ and $w_3 = 3$ are the weight values of y_1, y_2 and y_3, respectively.

4.4 Experimental Results

We trained the data set several times and used 60 hidden neurons. The NAR neural network good fitting generated a nonlinear curve for the 2011 sales data. See the graph below. We repeated this method for year 2012 and 2014. Then, we obtain three NAR neural networks based on three groups of training sets, 2011, 2012 and 2014, respectively.

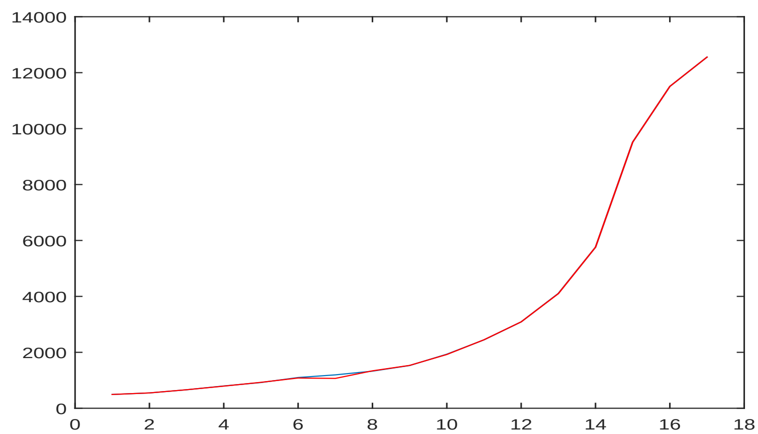

Figure 11: Fitting Curve of 2011, $M = 60$

A part of 10-step forecast results based on three different NAR neural networks one for each year are shown in the table below.

Table 4: 10-step Forecast Based on Three NAR Networks, $M = 60$

NAR	1	2	3	4	5	6	7	8	9	10
2011	586.10	712.84	874.64	1063.484	1171.81	1278.87	1358.84	1498.22	1952.05	2476.97
2012	618.76	812.3	436.99	69.63	125.49	1081.97	2412.80	6965.56	3786.26	6515.24
2014	461.83	486.34	523.25	520.44	533.90	583.85	613.2142	640.84	706.51	784.342

We calculated the weighted average and the final predictive value of the next ten weeks. The analysis of the results are shown in the figures.

Table 5: Final Predictive Values, $M = 60$

t	1	2	3	4	5	6	7	8	9	10
$y(\hat{t})$	584.85	632.74	553.07	460.68	504.08	865.73	1337.35	2891.974	1940.679	2976.74

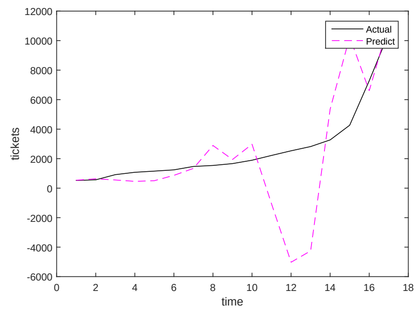

Figure 12: Results of Prediction, $M = 60$

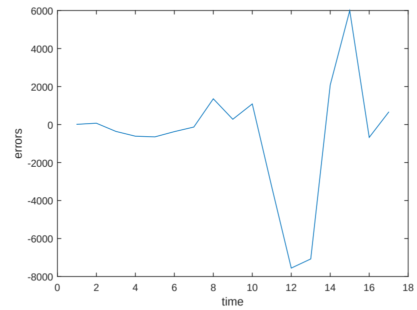

Figure 13: Error Between Actual And Predictive Values

When $t < 10$, the errors are small, that is, the prediction is good. However, when $10 < t < 16$, there is a huge error and this model fails. According to the principle of neural networks, the failure is due to over-fitting [11] with train datasets. When we changed $M = 60$ into a smaller number $M = 10$, and repeated the whole prediction process again, we get much better results. The fitted results of 2011 when $M = 10$ are shown in the figures.

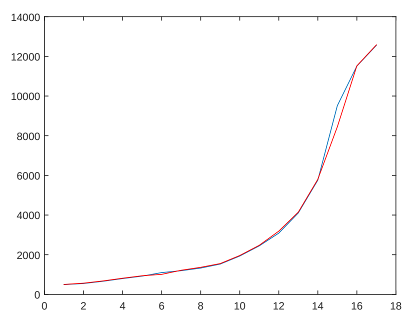

Figure 14: Fitting Curve of 2011, $M = 10$

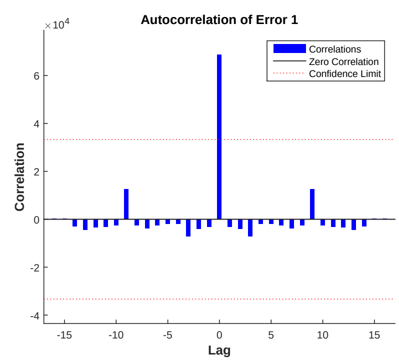

Figure 15: Autocorrelation of Error 2011, $M = 10$

We can see there is much more errors between the fitting data and the actual data when $M = 10$. The final prediction is the weighted average of predictive values of the three different NAR neural networks for year 2011, 2012 and 2014.

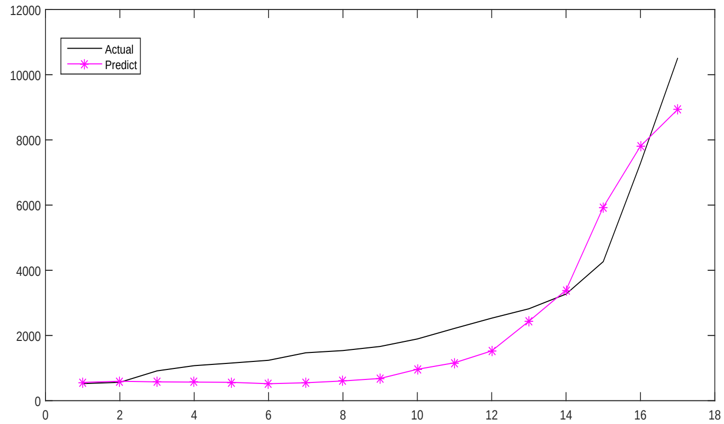

Figure 16: Actual And Predictive Values of Tickets Selling in 2016

4.5 Summary

In this chapter, we used the dynamic NAR neural network prediction model, which is well suited for taking into account. If any unexpected events occurs during ticket selling time. By using in-going ticket sales data, this method constantly updates the prediction and reduces the impact of unexpected events on the prediction. We presented the flow chart and the recursive equation underlying the structure and algorithm of the dynamic NAR neural network. Our results show that this method provides good short term prediction.

CHAPTER 5

CONCLUSIONS

To compare the results of the two prediction methods, the three-segment latent class Weibull model and the dynamic NAR neural network, respectively. We use the root mean squared error (RMSE) and the mean squared prediction error (MSPE).

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(y(t_i) - \hat{y}(t_i))^2} \tag{12}$$

$$MSPE = \frac{1}{K}\sqrt{\sum_{i=1}^{K}(\frac{y(t_i) - \hat{y}(t_i)}{y(t_i)})^2}. \tag{13}$$

Here K is the total number of data, $K = 26$; $y(t_i)$ is an array of actual values and $\hat{y}(t_i)$ is the corresponding predictive values. The evaluation results are provided in following table.

Table 6: Comparison Between LCW Model And DNN Model

	RMSE	MSPE
LCW	1027.9	0.0762
DNN	860.80	0.1002

Here the LCW represents the latent class Weibull model and the DNN represents the dynamic NAR neural network model. It can be seen that DNN has the lower RMSE value, but LCW has a lower MSPE.

The three-segment latent class Weibull model gives a better prediction for the upcoming event if a prediction is needed before ticket sales begin. However, the dynamic NAR neural network model is better suited for predicting ticket sales after sales have began, the dynamic NAR neural network model predicts short term prediction with higher accuracy.

BIBLIOGRAPHY

[1] Jacob Suher, *Forecasting Event Ticket Slaes*, Wharton Research Scholars, **49**(2008).

[2] Arthur J. Hallinan Jr., *A Review of the Weibull Distribution*, Journal of Quality Technology, **25**(1993), 85-93.

[3] Ali N. Celik, *A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey*, Renewable Energy, **29**(2004), 593-604.

[4] Jay Magidson, Jeroen K. Vermunt, *Latent Class Models*, Statistical Innovation Inc, (2004), 175-197.

[5] Khalil Benmouiza, Ali Cheknane, *Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models*, Energy Conversion and Management, **75**(2013), 561-569.

[6] Bruce G.S. Hardie, Peter S. Fader, *Applied Probability Models in Marketing Research*, Advanced Research Techniques Forum, **11**(2000), 569-575.

[7] Tsunghsien Tsai, Chikang Lee, Chienhung Wei, *Neural Network Based Temporal Feature Models for Short-term Railway Passenger Demand Forecasting*, Expert Systems with Applications, **36**(2009), 3728-3736.

[8] Zilin Gao, Jiaping Wang, Shuaitang Zhang, Xiuguo Zou, *Study of Plant Electrical Signal Based on Complex Wavelet and Dynamical Neural Network*, Journal of Nanjing Agricultural University, **40**(2017), 556-563.

[9] Jose Maria P. Menezes Jr., Guilherme A. Barreto, *Long-term Time Series Prediction with The NARX Network: An Empirical Evaluation*, Neurocomputing, **71**(2008), 3335-3343.

[10] Mauricio Mauledoux, Oscar Aviles, et al., *Analysis of Autoregressive Predictive Models and Artificial Neural Networks for Irradiance Estimation*, Indian Journal of Science and Technology, **9**(2016), 1-7.

[11] Nitish Srivastava, Geoffrey Hinton, et al., *A Simple Way to Prevent Neural Networks from Overfitting*,Journal of Machine Learning Research, **15**(2014), 1929-1958.