CHAPTER 3
IDENTIFICATION OF OUTLIER ON FUZZY TIME SERIES

3.1 INTRODUCTION

In this chapter presented Markov Model in Fuzzy Time Series (MM - FTS) based on outlier. As Song and Chissom [1993] pointed out, traditional forecasting methodologies are not suited to data composed of linguistic values. This is the motivation for the fuzzy time series method is introduced. There are probability models such as the discrete state Markov models, however, that use categorical data which can include linguistic labels. These are seldom applied in a traditional time series environment since data are most often available in numerical form. In this section a Markov model for the enrollment data will be described and the parameter estimation process compared with that of the fuzzy time series method.

A Markov chain is a discrete random process with the property that the next state depends only on the current state; the past states have no influence on the future. A Markov chain $X$ is said to be time-homogenous if the conditional probability

$$P[X_{n+1} = j | X_n = i] = P_{ij}, \ i, j \in S$$

(3.1)
is independent of \( n \) and \( S \) is the countable state space. The probabilities \( P_{ij} \) are then called the transition probabilities for the Markov chain \( X \); for any \( i, j \in S \), \( P_{ij} \geq 0 \), and \( \sum_{j \in S} P_{ij} = 1 \) for any \( m \in S \).

\[
P[X_{n+m} = j | X_n = i] = p_{ij}^m, \quad i, j \in S
\] (3.2)

Here \( P_{ij}^{(m)} \) denotes the probability that the process goes from state \( i \) to state \( j \) in ‘\( m \)’ transitions. The transition probabilities \( P_{ij} \) can be exhibited as a square matrix

\[
P = \begin{bmatrix}
P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\
P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\
P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots 
\end{bmatrix}
\] (3.3)

which is called as the transition probability matrix of the chain. If the number of states is finite, say \( n \), then there will be \( n \) rows and \( n \) columns in the matrix \( P \); otherwise the matrix will be infinite. As it is known, \( P_{ij} \geq 0 \), and \( \sum_{j=0}^{\infty} P_{ij} = 1 \) for every \( i, j = 0, 1, 2, \ldots \).

If \( P_{ij}^{(n)} \) denotes the probability that the process having started from the state \( i \) to reach the state \( j \) in \( n \) transitions, then the stationary probabilities \( \pi_0, \pi_1, \) and \( \pi_2 \) satisfy,
\[ \lim_{n \to \infty} P_{ji}^{(n)} \pi_j = \pi_i > 0, \text{ for all } j = 0,1,2,\ldots \text{ if} \]

\[ \lim_{n \to \infty} P_{ij}^{(n)} \pi_i = \pi_j > 0, \text{ for all } i = 0, 1, 2,\ldots \]

If for a given Stochastic Processes, \( P \) is the transition probability matrix, then the estimates of \( \pi_0, \pi_1, \) and \( \pi_2 \) are the steady state distributions of the system of equations.

\[ \pi P = \pi \text{ where } \pi = (\pi_0, \pi_1, \pi_2) \]

Even though the states of a Markov chain are mutually exclusive, the process being modeled does not have to occupy one particular state with certainty for a Markov model to be valid. Several states can have non-zero probabilities, analogous to the concept of fuzzy set membership. With the Markov model, however, there is the requirement that all the state probabilities must sum to one for each observation. Membership functions do not have this restriction.

Emulating the fuzzy time series model as closely as possible, we let each state in the Markov model correspond to a probability density function (pdf), referred to as \( b_i, i=1,\ldots,16. \) Each pdf is defined on the same universe, \( U, \) as the corresponding fuzzy set, having the same shape as the memberships of the intervals \( u_i \) in the corresponding fuzzy set, but scaled to unit area. In the Markov chain approach each state, analogous to a fuzzy set \( A, \) corresponds to a probability distribution on the universe of discourse instead of a fuzzy set. In addition, a year's observation will be a vector giving the
probabilities for each state instead of a vector giving memberships in the fuzzy sets. These probability distribution functions can be estimated in some applications or determined subjectively.

The fuzzy enrollment data will be used to estimate the transition matrix $R_m$ for a Markov model, using the same years [1971-1990] used by [Song and Chissom, 1993]. Each pair of successive years in the data constitutes a transition, so there are 21 observed transitions in the 22 years of enrollment data. Associated with each observed transition is a $p \times p$ matrix that is the outer vector product of the vector of probabilities corresponding to the initial observation times the vector of probabilities corresponding to the subsequent observation. Summing these matrices then normalizing all rows to sum to unity produces the estimated transition matrix for the Markov model. The forecast is this estimated transition matrix pre-multiplied by the previous year's observed (or forecasted) probability vector. Several variations in this technique are possible by using different probability distributions for the observations.

### 3.2 Definition of Outlier

Outlier defines an observation that is numerically distant from the rest of the data. Any observation in a set of data is inconsistent with remainder of the observations in the data set. The term ‘outliers’ is a little vague concept and has been viewed differently by different researchers. “An outlier in a set of data to be an observation
(or subset of observations) which appears to be inconsistent with the remainder of that set of data" – [Barnett and Lewis, 1984]. “An observation whose dependent variable value is conditionally unusual given the value of the independent variable” – [Fox, 1997].

Identification of outliers is a great task and it may lead to better performances in estimating predicted values. Outliers occur not only in chronological data but also in specified data which based on probability distribution. A perusal of the above descriptions makes it clear that most of them are merely vague descriptions of what an outlier is rather than precise definitions of the same. Because terms like ‘too small’, ‘too large’, ‘deviate markedly’ and ‘inconsistent’ are only relative in nature.

Suppose that in a sample of size 10, drawn from a standard normal population, one of the observations is 5.1. It is obvious that it is extremely unlikely to draw such an observations from a standard normal population. Yet it is possible that either an error has been committed (The actual observation may be 1.5 but wrongly recorded as 5.1) or this observations could have come from a normal population with a mean bigger than zero. Such an observation is considered as an outlier.

It is well known that statistical data, collected for analysis and interpretation, most often contains one or two measurements which
do not look similar to the rest of the data. Such measurements are vaguely called ‘outliers’.

They could have arisen naturally as rare events or due to human error in data collection or theoretical error in model selection. Unless these outliers are properly treated, it is possible that statistical conclusions based on such data are misleading. Hence, such ‘outliers’ should be identified and treated properly to draw proper conclusions from the data. The outliers are also known by the following names

- Discordant values.
- Rogue values.
- Contaminants.
- Surprising values.
- Mavericks.
- Dirty data.
- Fluky data.
- Unusual data
- Influential data.

The occurrence of outliers provides interesting case studies for further exploration. Their existence should be investigated and never be ignored. In any scientific research, full disclosure of data modeling is required, including a disclosure and discussion of outliers. There has been much debate in the literature regarding what to do with the existence of outliers in data sets including time series data. Studies had shown that outliers affect the performance of standard
statistical methodology in modeling, forecasting and diagnostic purposes. Different approaches of detecting and handling outliers have been considered with the objective of improving the efficiency and sufficiency of statistical analyses.

An obvious unrepresentative measurement error supports rejection of the offending observation. An outlier in the form of an excessive execution error may sometimes also lead to rejection. But it could on occasions warrant a modified model. Anscombe, F.J [1960] distinguishes in terminology between outliers arising from large variation of the inherent type, and these from large measurement or execution error. He calls the former ‘outliers’, the latter ‘spurious observation’. We shall make no such distinction. The full study of statistical methods for outliers needs to encompass all derivative sources of variation; the only exceptions are outliers arising from clearly discernible deterministic mistakes of calculation, recording, etc. in this case rejection is the only remedy; otherwise we need to clearly recognize the many possibilities, other than outright rejection, for coping with outliers.

### 3.3 Causes of Outliers

- Measurement error.
- Miscoding.
- Entered incorrectly.
- Misinterpretation.
- Lurking variable.
• Unusual points.
• Fluky data.

Figure 3.1: Treatment of Outliers

3.4 COOK’S DISTANCE

A measure of how much the residual of all cases would change if a particular case is excluded from the calculation of the regression coefficients. A large cook’s Distance [Dennis Cook, 2000] indicates
that excluding a case from computation of the regression statistics changes the coefficients substantially.

In Statistics, Cook’s distance is a commonly applied to estimate of the influence of a data point when doing least squares regression analysis. In an ordinary least squares analysis, Cook’s distance can be used in several ways

- To indicate data points that is particularly worth checking for validity.
- To indicate regions of the design space where it would be good to be able obtains more data points.

Cook’s distance measures the effect of deleting given observations. Data points with large residuals (outliers) and or high leverage may distort the outcome and accuracy of a regression. Points with a large Cook’s distance are considered to merit closer examination in the analysis.

\[
D_i = \sum_{j=1}^{n} \frac{(\hat{Y}_j - \hat{Y}_{j(i)})^2}{pMSE} \tag{3.4}
\]

The following is an algebraically equivalent expression
\[ D_i = \frac{e_i^2}{p \text{MSE}} \left[ \frac{h_{ii}}{(1-h_{ii})^2} \right] \]  

(3.5)

where

\( \hat{Y}_j \) - is the prediction from the full regression model for observation j.

\( \hat{Y}_{j(i)} \) - is the prediction for observation j from a refitted regression model in which observation i has been omitted.

\( h_{ii} \) - is the i\textsuperscript{th} diagonal element of the hat matrix \( X(X^TX)^{-1}X^T \).

\( e_i \) - is the crude residual (i.e., the difference between the observed and the fitted value by the proposed model). MSE is the mean square error of the regression model. p is the number of fitted parameters in the model.

There are different opinions regarding what cut-off values to use for spotting outliers. A simple operational guideline of \( D_i > 1 \) has been suggested. Some researchers indicated that \( D_i > 4/n \), where n is the number of observations.

### 3.5 STUDENTIZED RESIDUAL TEST

A Studentized residual is the quotient resulting from division of a residual by an estimate of its standard deviation. Typically the standard deviations of residuals in a sample vary greatly from one data point to another even when the errors have the same standard deviation, particularly in regression analysis [Paul et.al., 1991].
Thus it does not make sense to compare residuals at different data points without first studentizing. It is a form of Student’s t-statistic, with the estimate of error varying between points. Thus the studentized residual can be interpreted as a z-score. Approx. 95% of the Studentized residual values should be lie in -2 to 2. Extreme studentized residual values can indicate that a given observation is an outlier. This is an important technique in the detection of outliers (High Studentized residual (|Residual|>2)). It is named in honor of William Sealey Gosset, who wrote under the pseudonym Student, and dividing by an estimate of scale is called Studentizing, in analogy with standardizing and normalizing.

The Studentized residual analysis method can assist us to determine whether there exist outliers in the chronological data. The studentized test can be employed to examine the outliers. It described as follows:

If there are n chronological data $x_1, x_2, ..., x_n$ a square matrix $H$ can be defined as follows,

$$H = X(X^TX)^{-1}X^T$$
\[
X = \begin{bmatrix}
1 & X_1 \\
1 & X_2 \\
. & . \\
. & . \\
. & . \\
1 & X_n
\end{bmatrix}
\]

The Studentized residual can be defined by,

\[
\text{Studentized Residual Test} = \frac{e_i}{S_j}, 
\]

(3.7)

Where \( S_j = \hat{\sigma}(i)\sqrt{1-h_{ii}} \)

Here, \( S_j \) is the estimated variance of the residual, \( e_i \) specifies the residual of the \( i^{th} \) datum. \( \hat{\sigma}(i) \) is the estimated value of the standard deviation \( \sigma \) without the \( i^{th} \) observation, \( h_i \) is the \( i^{th} \) diagonal element of matrix H. The data is considered to be an outlier where the absolute residual values having studentized residuals greater than 2.0.

3.6 MM-FTS MODEL DESCRIPTION

A significant drawback of the fuzzy time series methods are high computational overheads owing to complex matrix operations and lower accuracy. The MM-FTS work aims to give a better forecasting accuracy using fuzzy time series and it is emphasized that the forecast uses only historical data. The significance of the MMFTS work is that it reduces the average forecasting error when compared with the existing forecast approaches like [Chen, 2002, Lee et al.,2004, Shivaraj Singh, 2007].
The step by step forecasting procedure as follows:

**Step 1.** First identifying outliers from the historical data using Cook’s distance and Studentized residual test.

**Step 2.** After identifying outlier, compute the appropriate length of interval \( l \) using distribution based method.

**Step 3.** Compute the number of intervals \( m \) as follows:

\[
m = \frac{(V_{\text{max}} + V_2) - (V_{\text{min}} - V_1)}{l}
\]  

(3.8)

where \( V_{\text{max}} \) is the maximum value of the historical data, \( V_2 \) is the positive integer, \( V_{\text{min}} \) is the minimum value of the historical data, \( V_1 \) is the positive integer and \( l \) is the appropriate length of interval computed in step (2).

**Step 4.** Let \( U \) be the universal set,

\[
U = [V_{\text{min}} - V_1, V_{\text{max}} + V_2]
\]

and partition \( U \) into equal length of intervals \( \{u_1, u_2, u_3 \ldots u_m\} \).

**Step 5.** Fuzzify the variations of the historical data and determine Fuzzy logical relationships.
Step 6. If $A_i$ is fuzzify the value of current year $n$ and $A_j$ is the fuzzify the value of next year $n+1$, then fuzzy logical relation is denoted by $A_i \rightarrow A_j$.

Step 7. Define Fuzzy sets $A_i$ on universe of discourse $U$. Then determined how many linguistic variables to be fuzzy sets.

Step 8. Define the linguistic terms of $A_i$ represented by the fuzzy sets as follows:

- $A_1 = \{u_1/0.667, u_2/0.337, u_3/0, \ldots, u_m/0\}$
- $A_2 = \{u_1/0.25, u_2/0.5, u_3/0.25, \ldots, u_m/0\}$
- $A_3 = \{u_1/0, u_2/0.25, u_3/0.5, \ldots, u_m/0\}$

...  

- $A_m = \{u_1/0, \ldots, u_{m-1}/0.333, u_m/0.667\}$

Step 9. Fuzzify the historical data as follows: If the value belongs to $u_1$, then fuzzified membership values into $0.667/A_1+0.333/A_2+0/A_3$ denoted by $A_1$. If the value belongs to $u_i$, $i = 2,3, \ldots, n-1$, then the fuzzified membership values into $0.25/A_{i-1}+0.5/A_i+0/A_{i+1}$ denoted by $A_i$. If the value belongs to $u_n$ then the fuzzified membership values $0/A_{n-2}+0.333/A_{n-1}+0.667/A_n$ denoted by $A_n$.

Step 10. Identify the fuzzy logical relationship first order fuzzy time series as follows,

$$A_{j-1} \rightarrow A_j.$$
Step 11. Determine logical relationship

\[ R_i = A_{i-1}^T \times A_i, \ i = 1, 2, \ldots, n \quad (3.9) \]

and obtain the relation matrix \( R \) as follows:

\[ R = \bigcup_{i=1}^{n} R_i \quad (3.10) \]

Step 12. Obtain the transition probability matrix from the fuzzy relation matrix \( R \). Calculate the estimated values using state transitions probability vector membership function as follows.

\[ P_{t+1}' = P_t' \times P_m', \quad (3.11) \]

where, \( P_{t+1}' \) is the current year historical data is obtained from previous year vector probability membership \( P_t' \) and probability matrix \( P_m' \).

Step 13. Finally obtain the average forecasting error using actual and forecasted values.

\[ \text{Forecast error} = \frac{|\text{forecasted value} - \text{actual value}|}{\text{actual value}} \times 100\% \quad (3.12) \]

\[ \text{Average Forecast Error} = \frac{\text{sum of the forecasting errors}}{\text{total number of errors}}. \]
3.7 COMPUTATION RESULTS

The MM - FTS approach be explained with actual data corresponding to the number of accident occurred in India. The methodology is explained with original data set as follows:

From table 3.1, the unusual residual value -2.597 in the year 1992 which have Studentized residuals greater than 2.0 in absolute value. Studentized residuals measure how many standard deviations each observed value deviates from a model fitted using all of the data except that observation. In this case, there is one Studentized residual greater than 2.0, but none greater than 3.0. The step by step procedure as follows:

Table 3.1: Accident Data during 1985 - 2007

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Accident</th>
<th>Fuzzy Set</th>
<th>Cooks Distance</th>
<th>Student Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>20700</td>
<td>A₁</td>
<td>0.001</td>
<td>-0.168</td>
</tr>
<tr>
<td>1986</td>
<td>21550</td>
<td>A₁</td>
<td>0.008</td>
<td>-0.415</td>
</tr>
<tr>
<td>1987</td>
<td>23400</td>
<td>A₂</td>
<td>0.000</td>
<td>0.026</td>
</tr>
<tr>
<td>1988</td>
<td>24670</td>
<td>A₃</td>
<td>0.000</td>
<td>0.067</td>
</tr>
<tr>
<td>1989</td>
<td>27000</td>
<td>A₄</td>
<td>0.020</td>
<td>0.837</td>
</tr>
<tr>
<td>1990</td>
<td>28260</td>
<td>A₅</td>
<td>0.019</td>
<td>0.870</td>
</tr>
<tr>
<td>1991</td>
<td>29340</td>
<td>A₅</td>
<td>0.014</td>
<td>0.777</td>
</tr>
<tr>
<td>1992</td>
<td>26030</td>
<td>A₄</td>
<td><strong>0.115</strong></td>
<td><strong>-2.597</strong></td>
</tr>
<tr>
<td>1993</td>
<td>28010</td>
<td>A₅</td>
<td>0.067</td>
<td>-1.890</td>
</tr>
<tr>
<td>1994</td>
<td>32040</td>
<td>A₇</td>
<td>0.000</td>
<td>0.142</td>
</tr>
<tr>
<td>1995</td>
<td>34890</td>
<td>A₈</td>
<td>0.037</td>
<td>1.292</td>
</tr>
<tr>
<td>1996</td>
<td>37120</td>
<td>A₉</td>
<td>0.097</td>
<td>2.126</td>
</tr>
<tr>
<td>1997</td>
<td>37370</td>
<td>A₉</td>
<td>0.048</td>
<td>1.345</td>
</tr>
<tr>
<td>1998</td>
<td>38500</td>
<td>A₁₀</td>
<td>0.051</td>
<td>1.290</td>
</tr>
<tr>
<td>Year</td>
<td>Value</td>
<td>Interval</td>
<td>Width</td>
<td>Area</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>1999</td>
<td>38640</td>
<td>A₁₀</td>
<td>0.010</td>
<td>0.524</td>
</tr>
<tr>
<td>2000</td>
<td>39140</td>
<td>A₁₀</td>
<td>0.000</td>
<td>0.037</td>
</tr>
<tr>
<td>2001</td>
<td>40560</td>
<td>A₁₁</td>
<td>0.002</td>
<td>0.182</td>
</tr>
<tr>
<td>2002</td>
<td>40750</td>
<td>A₁₁</td>
<td>0.016</td>
<td>-0.528</td>
</tr>
<tr>
<td>2003</td>
<td>40670</td>
<td>A₁₁</td>
<td>0.040</td>
<td>-1.504</td>
</tr>
<tr>
<td>2004</td>
<td>42990</td>
<td>A₁₂</td>
<td>0.035</td>
<td>-0.667</td>
</tr>
<tr>
<td>2005</td>
<td>43920</td>
<td>A₁₂</td>
<td>0.070</td>
<td>-0.882</td>
</tr>
<tr>
<td>2006</td>
<td>46090</td>
<td>A₁₄</td>
<td>0.0061</td>
<td>-0.0852</td>
</tr>
<tr>
<td>2007</td>
<td>47920</td>
<td>A₁₄</td>
<td>0.1347</td>
<td>0.37187</td>
</tr>
</tbody>
</table>

First the appropriate length of interval $l$ is computed using distribution based length procedure as follows. Finally the length of interval $l = 2000$.

The calculated number of intervals $m$ as follows,

$$m = \frac{48000 - 20000}{2000} = 14.$$  

From table 3.1, we get $V_{\text{min}} = 20700$ and $V_{\text{max}} = 47920$. The variable $V₁$ and $V₂$ are two positive real numbers, properly chosen by the user. Suppose we get $V₁=700$ and $V₂=80$ then define universe of discourse, $U = [20000, 48000]$, and $U$ is partitioned into 14 equal length of intervals $u_i$, $i = 1, 2, \ldots, 14$,

$$u₁ = [20000, 22000),$$
$$u₂ = [22000, 24000),$$
$$u₃ = [24000, 26000),$$
$$\ldots,$$
\[ u_{13} = [44000, 46000), \]
\[ u_{14} = [46000, 48000]. \]

It is assumed that the linguistic variable of the historical data can take fuzzy values are as follows:

- \( A_1 \) - (not many),
- \( A_2 \) - (not too many),
- \( \ldots \),
- \( A_{13} \) - (too many) and
- \( A_{14} \) - (too many many)

Then for the given intervals \( u_i \), \( i = 1, 2, \ldots, 14 \), each \( u_i \) belongs to a particular \( A_j \), \( j = 1, 2, \ldots, 14 \) and is expressed by the real value within the range \([0,1]\).

Fuzzification is the process of identifying associations between the historical values in the data set and the fuzzy sets defined in the previous step. If the maximum degree belonging to certain time variable, say \( F(t-1) \), occurs at fuzzy set \( A_j \), then \( F(t-1) \) is fuzzified as \( A_j \).

According to table 3.1, the number of accident in the year 2000 was 39140 which lies in the fuzzy interval \( A_{10} \). Since the maximum membership degree of \( U_{10} \) occur at \( A_{10} \), the historical data \( F(2000) \) is fuzzified as \( A_{10} \).
Identify the first order fuzzy logical relationship of the historical data. From table 3.1, we can see that year 1999 and 2000 both are fuzzified as $A_{10}$, which provides the relationship as follows, $A_{10} \rightarrow A_{10}$.

The complete set of relationship are shown in table 3.2,

Table 3.2: First order fuzzy logical relationship

<table>
<thead>
<tr>
<th>$A_1 \rightarrow A_1$</th>
<th>$A_1 \rightarrow A_2$</th>
<th>$A_2 \rightarrow A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_3 \rightarrow A_4$</td>
<td>$A_4 \rightarrow A_5$</td>
<td>$A_5 \rightarrow A_5$</td>
</tr>
<tr>
<td>$A_5 \rightarrow A_4$</td>
<td>$A_4 \rightarrow A_5$</td>
<td>$A_5 \rightarrow A_7$</td>
</tr>
<tr>
<td>$A_7 \rightarrow A_8$</td>
<td>$A_8 \rightarrow A_9$</td>
<td>$A_9 \rightarrow A_9$</td>
</tr>
<tr>
<td>$A_9 \rightarrow A_{10}$</td>
<td>$A_{10} \rightarrow A_{10}$</td>
<td>$A_{10} \rightarrow A_{10}$</td>
</tr>
<tr>
<td>$A_{10} \rightarrow A_{11}$</td>
<td>$A_{11} \rightarrow A_{11}$</td>
<td>$A_{11} \rightarrow A_{11}$</td>
</tr>
<tr>
<td>$A_{11} \rightarrow A_{12}$</td>
<td>$A_{12} \rightarrow A_{12}$</td>
<td>$A_{12} \rightarrow A_{14}$</td>
</tr>
<tr>
<td>$A_{14} \rightarrow A_{14}$</td>
<td>$A_{13} \rightarrow A_{14}$</td>
<td>$A_{14} \rightarrow A_{14}$</td>
</tr>
<tr>
<td>$A_{14} \rightarrow A_{14}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The fuzzy relationships are combined into fuzzy logical relation starting from identical left-hand sides. Then $R_i, i = 1, 2...14$, is calculated as a sum of logical relationships in each group. Here, the relation matrix $R_i$ is converted into transition probability matrix $P_m$, which is shown in below,

$$ R_i = A_i^T \times A_i $$
\[ R = \bigcup_{i=1}^{n} R_i \]

\[
\begin{bmatrix}
0.67 & 0.74 & 0.35 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.43 & 0.71 & 0.65 & 0.31 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.06 & 0.31 & 0.56 & 0.25 & 0.25 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.06 & 0.19 & 0.31 & 0.38 & 0.25 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.06 & 0.25 & 0.44 & 0.38 & 0.13 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.06 & 0.38 & 0.63 & 0.44 & 0.19 & 0.06 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.25 & 0.44 & 0.50 & 0.38 & 0.13 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.13 & 0.25 & 0.19 & 0.19 & 0.25 & 0.19 & 0.06 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.06 & 0.13 & 0.06 & 0.00 & 0.13 & 0.31 & 0.25 & 0.06 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.06 & 0.31 & 0.50 & 0.31 & 0.06 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.19 & 0.63 & 0.75 & 0.38 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.06 & 0.44 & 0.94 & 0.90 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.13 & 0.56 & 1.03 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.13 & 0.56 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.74
\]
The above table represents the defuzzified forecast outputs using transitions state probability membership function. The outputs are multiplied with corresponding mid values of the fuzzy interval over the period of years and its overall summation leads the predicted values. For example, year 2004 is forecasted using fuzzified values of 2003. The midpoints of the intervals $U_1, U_2, \ldots, U_{14}$ are multiplied into corresponding defuzzified probability values and its overall summation. The actual and predicted value of number of accidents in India is shown in figure 3.2.

Finally, the average forecasting error was obtained using actual and estimated values, when compared with the other existing fuzzy time series methods. The result is shown in table 3.4.
<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
<th>Predicted Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>20700</td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>21550</td>
<td>23202</td>
</tr>
<tr>
<td>1987</td>
<td>23400</td>
<td>23202</td>
</tr>
<tr>
<td>1988</td>
<td>24670</td>
<td>24488</td>
</tr>
<tr>
<td>1989</td>
<td>27000</td>
<td>26858</td>
</tr>
<tr>
<td>1990</td>
<td>28260</td>
<td>28952</td>
</tr>
<tr>
<td>1991</td>
<td>29340</td>
<td>30280</td>
</tr>
<tr>
<td>1992</td>
<td>30400</td>
<td>30280</td>
</tr>
<tr>
<td>1993</td>
<td>32040</td>
<td>30280</td>
</tr>
<tr>
<td>1994</td>
<td>34890</td>
<td>34958</td>
</tr>
<tr>
<td>1995</td>
<td>37120</td>
<td>37042</td>
</tr>
<tr>
<td>1996</td>
<td>37730</td>
<td>38477</td>
</tr>
<tr>
<td>1997</td>
<td>38500</td>
<td>38477</td>
</tr>
<tr>
<td>1998</td>
<td>38640</td>
<td>39996</td>
</tr>
<tr>
<td>1999</td>
<td>39140</td>
<td>39996</td>
</tr>
<tr>
<td>2000</td>
<td>40560</td>
<td>39996</td>
</tr>
<tr>
<td>2001</td>
<td>40750</td>
<td>41656</td>
</tr>
<tr>
<td>2002</td>
<td>40670</td>
<td>41656</td>
</tr>
<tr>
<td>2003</td>
<td>42990</td>
<td>41656</td>
</tr>
<tr>
<td>2004</td>
<td>43920</td>
<td>43441</td>
</tr>
<tr>
<td>2005</td>
<td>46090</td>
<td>43441</td>
</tr>
<tr>
<td>2006</td>
<td>47920</td>
<td>46001</td>
</tr>
</tbody>
</table>

Table 3.3: Forecasting number of accident from 1985-2007
Figure 3.2: Actual and Predicted values of accidents in India

Table 3.4: Average Forecasting Error of Fuzzy Time series methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AFE</td>
<td>3.90%</td>
<td>3.43%</td>
<td>2.89%</td>
<td>2.60%</td>
</tr>
</tbody>
</table>
3.8 CONCLUSION

Time series analysis is often associated with the discovery of patterns and prediction of features. The fuzzy time series has recently attracted more attention because of its capability of dealing with vague and incomplete data. There have been a variety of models developed to either improve forecasting accuracy or reduce computation overhead. In this chapter, we proposed MM-FTS method is mainly focusing on improving the forecasting accuracy by removing the identified outliers in the data set. Outlier plays a vital role in time series data; it is mainly affecting the prediction accuracy, so first identified outlier from the time series data using cook’s distance and studentized residual test. After MM-FTS method is adopted for the data related to
accidents data in India. The experiment result found that the average forecasting error is 2.86% before removing the outliers. After removing the outlier, the method produces 2.6% of average forecasting error. Thereby, the MM - FTS method improves average forecasting accuracy by around 9%. The results indicates that MM –FTS method is more appropriate compared with other existing FTS methods. It supports numerical and graphical representations.