Exploring finite mixtures of Gaussian Innovations in ARX(1) Model

Dr. Anuj Nain[1*]

[1]Assistant Professor, Department of Decision Science, New Delhi Institute of Management, New Delhi, India | Email: anuj.nain@ndimdelhi.org

***Corresponding Author:** Dr. Anuj Nain | Email: anuj.nain@ndimdelhi.org

Abstract

Researchers have introduced and studied AR(1) model with explanatory variable to model a variety of time series data. A usual assumption in these models is that the innovations follow a single gaussian distribution. There might occur situations in which errors in the model may arise from several subpopulations rather than a single density. In this study a mixture distribution to model the innovations in an AR(1) Model with Explanatory variable has been introduced. A simulation study has been conducted to justify the proposed theory. An Empirical analysis on real time data set has also been conducted to provide application.

Keywords; AR(1) model with explanatory variable, Mixture density, EM Algorithm, Innovations, White Noise.

1. Introduction and Literature review

Time series modelling is a powerful and indispensable tool in the realm of data analysis, providing a systematic approach to understanding and forecasting trends over time. Drawing from the theories of time series models presented by Brockwell and Davis (1991), Box et.al (2015), Shumway et.al (2000) and Hamilton (2020), It is a statistical technique that enables the analysis of temporal data points, offering insights into patterns, fluctuations, and dependencies within a sequence. As businesses, researchers, and decision-makers grapple with an increasing influx of time stamped data, the significance of time series modeling becomes more pronounced. The time series modelling as a statistical approach is used to analyze and make predictions about data points over time. In a time series, observations are recorded sequentially at equally spaced intervals, and the order of the observations is critical. Common examples of time series data include stock prices, weather patterns, economic indicators, fertilizer consumption, daily Box-office collection and daily sales figures etc,. The main objectives of time series modeling are to understand the underlying patterns and trends within the data and to make predictions or forecasts for future time points. Time series models help capture and analyse the temporal dependencies present in the data.

There are various types of models available in time series, such as AR, MA, ARMA, ARIMA, ARCH, GARCH etc. This research work focuses on Autoregressive (AR) models. The Autoregressive models use past observations as input to predict current value. These models are commonly abbreviated as AR(p) models where "p" is the number of lagged observations used in the model also referred as the order of the autoregressive process. Without loss of generality, the term " AR(p) models" refers to linear AR(p) models, however theoretically the Non-Linear version also exists. Some notable references on the discussion includes Jones and Cox (1978), Tjøstheim (1990), Zakoian and Howell (1992), An and Huang (1996), Fan and Yao (2003), Rao et.al (2012), Douc et. al(2014) , Chatfield (2019), and Balakrishna (2022),

For p=1 the AR(p) model becomes AR(1) model. An AR(1) model indicates a first-order autoregressive model, which relies on the immediately preceding observation to predict the current value. AR(1) models are quite popular and widely used in time series analysis. In this paper, an AR(1) model with explanatory variable ,(A popular version of AR(1) model) has been explored and utilized for

modelling the time series data. Researchers have introduced and studied AR(1) model with explanatory variable to model a variety of time series data. An AR(1) model with an explanatory variable [Box et.al (2015) and Hamilton (2020)] extends the basic autoregressive framework by incorporating additional predictor variables to improve the model's explanatory power and accuracy. In many real-world scenarios, the behavior of a time series can be influenced by additional factors or variables, which the basic AR(1) model does not account for. To address this limitation, the AR(1) model can be extended to include one or more explanatory variables. This results in a more flexible and informative model, capturing the effects of these additional factors on the time series.

Many times in real-life situations the random variables are not generated from a single but from a mixture of several distributions. In such scenario a single density is not adequate to represent the entire data, instead several subpopulations, each with its own probability distribution, are considered to model the data more accurately. This approach leads to the use of mixture models, where each subpopulation corresponds to a component in the mixture, allowing for better representation of the overall variability in the data. The early work on mixture distribution started with finite mixtures of Gaussian distribution. One of the earliest references is Pearson's work on the classification of skewed frequency curves. In his seminal paper "*Contributions to the Mathematical Theory of Evolution*", Pearson (1894) introduced the concept of fitting mixture distributions to biological data using a mixture of Gaussian distributions. In the mid-20th century, Anderson(1958) contributed to the theoretical understanding of mixture distributions in his paper "*An Introduction to Multivariate Statistical Analysis*", discussing mixture models in the context of multivariate data. Over the years, several researchers have contributed to the development of the theory of mixture distributions to model various types of data. However, since the specific applications of their work are not relevant to our study, we will focus solely on the basic concepts of mixture distributions. Next, we define the concept of mixture distributions, with the definitions being based on the theories presented by Chandra (1977), Everitt and Hand (1981), Redner and Walker (1984), Everitt (1996), Everitt (2013) and Miljkovic and Grün (2016)]. The mixture distribution is a weighted sum of K distributions where the weights sum up to one. Each weight corresponds to the proportion or contribution of the corresponding component density. Each density in the mixture is characterized by an unknown parameter (or a parameter vector).

A usual assumption in AR(1) model with Explanatory variable is that the error term follows a single density, However in time series modelling, we often encounter scenarios where the error (or innovation) terms arise from multiple subpopulations or clusters. Each cluster has its own weight, representing its contribution as a proportion to the overall error distribution. In These situations, a single density can be used to model the innovations rather a mixture density is appropriate. Since mixture distributions are well used for capturing humps or multimodality in the data that provides relevant descriptions about the underlying clusters,

this study proposes and explores finite mixture distribution as Innovations in AR(1) model with explanatory variable. The incorporation of a finite mixture of Gaussian distributions for the error terms introduces a sophisticated level of complexity and realism to the model, allowing for the capture of more intricate patterns and dependencies in time series data. By thoroughly examining the theoretical underpinnings, model specifications, and practical implications of this enhanced AR(1) model, we aim to provide a comprehensive understanding of its potential advantages and applications. This detailed exploration highlights the innovative nature of combining explanatory variables with mixture-distributed innovations, illustrating how this approach can lead to more accurate and insightful analyses. The applicability of the model has been discussed through simulation study and empirical analysis, showcasing its effectiveness in capturing the clusters.

The chronology of the work presented in the paper is as follows: In section 2 we discuss the Modeling and Estimation, section 2.1 gives a brief introduction about the Mixture distributions, section 2.2 presents the model under study an its extension incorporating the mixture density. Section 2.3 discusses about parameter estimation using EM Algorithm. Section 3 presents a simulation study to

support the proposed theory followed by empirical analysis in section 4 to show a real-time data application of the model. Section 5 presents the conclusion of the work.

2 Modeling and estimation

2.1 Mixture Distributions

Let X_1 X_2 . . . X_n be independent and identically distributed random sample from K-component finite mixture of probability distributions see Chandra (1977), Everitt(1996, 2013), Miljkovic and Grün (2016)). This mixture distribution is represented as

$$f(x;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(x;\theta_K), \quad \text{subject to } \sum_{k=1}^{K} \pi_k = 1 \qquad (1)$$

where $\boldsymbol{\theta} = (\pi', \theta') = (\pi_1, \pi_2, \ldots, \pi_{k-1}, \theta_1, \theta_2, \ldots \theta_k)$ is the vector of unknown parameters and $0 < \pi_i \leq 1$. These K distributions may or may not be from the same family. In the present study we assume that for the mixture density given in (1) the component densities $f_k(.)$ are from the same family. Further we implement EM algorithm for parameter estimation. In this paper $f_k(.)$ in (1) follows $N(0, \sigma_k^2)$. An "m" component Normal mixture distribution would be represented as m − K $N(0, \sigma_k^2)$ where k = 1, 2 ... m.

2.2 The Model Under Study

The AR(1) model with explanatory variable under study is represented as

$$y_t = \rho y_{t-1} + \alpha X_t + \varepsilon_t \qquad t = 1, 2 \ldots\ldots n \qquad (2)$$
$$or\ \varepsilon_t = y_t - \rho y_{t-1} - \alpha X_t \qquad t = 1, 2 \ldots\ldots n$$

Where X_t is the explanatory variable with $-\infty < \alpha < \infty$,

$$|\rho| < 1, \text{and } \varepsilon_t \sim m - K\, N(0, \sigma_k^2)$$
$$k = 1\,2, \ldots m$$

$N(0, \sigma_k^2)$ represents Gaussian distribution with mean 0 and variance σ_k^2. So that the pdf associated with k^{th} component is given by

$$f_k(\varepsilon_t;\ \sigma_k) = \begin{cases} \dfrac{e^{-\left(\frac{\varepsilon_t}{\sigma_k}\right)^2}}{\sqrt{2\pi}\sigma_k} & -\infty < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

It is to be noted that X_t can be either a time series or independent variables. If X_t is a time series, it means that the explanatory variable also has a temporal structure and its values are recorded over time. For example, if y_t represents the sales of a product, might represent the advertising expenditure for the same product over the same period. In this case,would influence while also having its own temporal dynamics. Similarly, X_t can also be independent variables, meaning they do not necessarily have a temporal structure. These could be variables that influence y_t but are not inherently time-dependent. For example, X_t could represent categorical factors such as the presence of a promotional event (yes/no) or other external conditions that can affect can be either a time series or independent variables. In this study X_t are independent variables.

2.3 Parameter estimation

The Expectation-Maximization (EM) algorithm is a statistical technique used for estimating the parameters of a probabilistic model when the data involves latent (unobserved) variables. It is particularly useful in situations where there is incomplete or missing information. This algorithm has garnered significant interest following the influential work of Dempster et.al (1977). The algorithm stands out as a versatile method for iteratively computing maximum likelihood estimates, particularly when observations can be perceived as incomplete data. For a comprehensive understanding of the theory behind the EM algorithm and its extensions, refer to Dempster et.al (1977), McLachlan and Krishnan (2007).

The White Noise process described for the model in section (2.2) assumes that clusters are the source of the Innovations, enabling us to represent them using a mixture density. Here we assume that the

entire set of innovations contains observed terms and Latent variables. By employing the EM Algorithm, we can attain maximum likelihood estimates when handling this mixture density. We call the density given in (1) as the incomplete data density. with the following likelihood function and log likelihood functions respectively.

$$\boldsymbol{l\,(\theta;\,X)} = \prod_{i=1}^{n} f\,(x_i;\,\boldsymbol{\theta}) = \prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k f_k\left(\varepsilon_i;\,\theta_k\right) \quad (3)$$

$$\boldsymbol{L\,(\theta;\,X)} = \sum_{i=1}^{n} log f\,(\varepsilon_i;\,\boldsymbol{\theta}) \quad (4)$$

We call the likelihood given in (4) as incomplete data log-likelihood whereas for implementing EM Algorithm, complete data log likelihood is required. So, at first we assume that the complete data-set consists of **D = (E, Z)** but that only **E** is observed , whereas **Z** = ($z_{ik} \in \{0, 1\}$, k = 1 , 2 . . . K , i = 1,2, . . .n) is the set of latent variables. The complete-data likelihood is then denoted by

$$\boldsymbol{l(\theta;\,E\,,\,Z)} = \prod_{i=1}^{n}\sum_{k=1}^{K}(\pi_k f_k(\varepsilon_i;\,\theta_k))^{z_{ik}} \quad (5)$$

where z_{ik} = 1 if the observation ε_i comes from kth component density, otherwise z_{ik} = 0. Taking logarithm of (7) gives us complete data log-likelihood.

$$\boldsymbol{L\,(\theta;\,X,\,Z)} = \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}[\log(\pi_k) + \log(f_k\left(\varepsilon_i;\,\theta_k\right))] \quad (6)$$

where $\boldsymbol{\theta}$ is the unknown parameter vector for which we wish to find the MLE. The term EM stands for Expectation – Maximization, which are also the two steps of the algorithm.

The E-step of the EM algorithm computes the expected value of $\boldsymbol{L\,(\theta;\,E,\,Z)}$ given the observed data, $\boldsymbol{E}$, and the current parameter estimate, $\boldsymbol{\theta}_{old}$ say. In particular, we define w_{ik} as

$$w_{ik} = \mathrm{E}\,[z_{ik} \mid \varepsilon_i,\,\boldsymbol{\theta}_{old}] = \frac{\pi_k f_k(x_i;\,\theta_k)}{\sum_{k=1}^{K}\pi_k f_k(\varepsilon_i;\,\theta_k)} \quad (7)$$

We replace z_{ik} by w_{ik}. in (9). The result is called as **Q** function , given by

$$\mathbf{Q}\,(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{old}) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}[\log(\pi_k) + \log(f_k\left(\varepsilon_i;\,\theta_k\right))] \quad (8)$$

In the M-step the **Q** function is maximized to obtain new estimates of $\boldsymbol{\theta}$ (θ and π) . The two steps are repeated as necessary until the sequence of new estimates of θ and π converges. This optimization problem can be solved separately for π and for θ. The estimates of π are updated in the sth iteration by

$$\pi_k^{(s)} = \frac{\sum_{i=1}^{n} w_{ik}^{(s)}}{n} \quad (9)$$

For σ_k the substitution leads to Eq.(10),which on simplification gives Eq.(11).

$$\sum_i w_{ik}\ \left[\frac{(y_i - y_i - \rho y_{i-1} - \alpha X_i)^2 - \sigma_k^2}{\sigma_k^3}\right] = 0 \quad (10)$$

$$\sigma_k = \sqrt{\frac{\sum_i w_{ik}(y_i - \rho y_{i-1} - \alpha X_i)^2}{\sum_i w_{ik}}} \quad (11)$$

For ρ the substitution leads to Eq.(12), which on simplification gives Eq.(13)

$$\sum_i \sum_k w_{ik}\left(\frac{y_{i-1}(y_i - \rho y_{i-1} - \alpha X_i)}{\sigma_k^2}\right) = 0 \quad (12)$$

$$\rho = \frac{\sum_i \sum_k w_{ik}\left(\frac{y_{i-1}(y_i - \rho y_{i-1} - \alpha X_i)}{\sigma_k^2}\right)}{\sum_i \sum_k w_{ik}\left(\frac{y_{i-1}^2}{\sigma_k^2}\right)} \quad (13)$$

Similarly for α the substitution leads to Eq.(14)

$$a = \frac{\sum_i \sum_k w_{ik}\left(\frac{X_i(y_i - \rho y_{i-1} - \alpha X_i)}{\sigma_k^2}\right)}{\sum_i \sum_k w_{ik}\left(\frac{X_i^2}{\sigma_k^2}\right)} \quad (14)$$

Eq.s (12), (13) and (14) can be solved to obtain the parameter estimates of σ_k ρ *and* a respectively. The estimates of π_k are obtained using Eq. (9). The E-Step and M-Step of the algorithm can summarized as follows:

E-Step: Update the weight w_{ik} *using Eq. (7)*

M-Step: Update the parameters using Eq.s (9) , (11) , (13)and (14)

Repeat the two steps until the estimates converge.

3 Simulation study

Simulation in statistical analysis involves generating artificial data using the model's assumptions and specifications. This artificial data undergoes the same statistical analysis as the actual data, allowing us to grasp the model's behavior under diverse dimensions, parameters, or conditions. To assess the proposed methodology, an extensive simulation study has been conducted , showcasing how effectively the model captures the clusters in the Innovation. In this section, a comprehensive simulation study, which demonstrates the effectiveness of the model in capturing the clustering nature of the innovations in time series data, has been presented. This evaluates the performance of our proposed methodology. For the model defined in (1), the innovations are driven by mixture density having " m " Gaussian components. For illustration, various cases with m = 2 (Model1) and m=3 (Model2) have been presented.

The AR(1) model with explanatory variable, with m=2 can be written as

$$y_t = \rho y_{t-1} + \alpha X_t + \varepsilon_t \qquad t = 1, 2 \ldots \ldots n$$

$$or\ \varepsilon_t = y_t - \rho y_{t-1} - \alpha X_t \qquad t = 1, 2 \ldots \ldots n$$

Where X_t is the explanatory variable distributed as *N(4,1)* $-\infty < a < \infty$,

$$|\rho| < 1 \text{ , and } \varepsilon_t \sim 2 - K\, N\left(0, \sigma_k^2\right), k = 1,2$$

There are six parameters in the model ($\rho, \alpha, \sigma_1, \pi_1, \sigma_2, and\ \pi_2$). Their estimation is done using EM Algorithm as discussed in section 2.3. The AR(1) model with explanatory variable, with m=3 can be written as

$$y_t = \rho y_{t-1} + \alpha X_t + \varepsilon_t \qquad t = 1, 2 \ldots \ldots n$$

$$or\ \varepsilon_t = y_t - \rho y_{t-1} - \alpha X_t \qquad t = 1, 2 \ldots \ldots n$$

Where X_t is the explanatory variable distributed as *N(4,1)* $-\infty < a < \infty$,

$$|\rho| < 1 \text{ , and } \varepsilon_t \sim 3 - K\, N\left(0, \sigma_3^2\right)\ k = 1,2,3$$

There are eight parameters in the model($\rho, \alpha, \sigma_1, \pi_1, \sigma_2, \pi_2, \sigma_3, and\ \pi_3$). Their estimation is done using EM Algorithm as discussed in section 2.3.

Simulation study for total of five different time series data sets have been conducted, which comprises of three series from Model1 and two from Model 2. At first, the parameters are initialized and then data is generated using these initial values. Then it is observed that the estimates lie close to the initial values (For example corresponding to Model 1 the initial values used to generate the first series are N= 300 , $\rho = 0.55,\ \alpha = 2, \sigma_1 = 3, \sigma_2 = 15, \pi_1 = 0.6,\ \pi_2 = 0.40$ and the parameter estimates obtained through EM Algorithm are $\hat{\rho} = 0.576,\ \hat{\alpha} = 1.878, \hat{\sigma_1} = 3.11,\ \widehat{\sigma_2} = 15.527, \widehat{\pi_1} = 0.578, \widehat{\pi_2} = 0.422$. It is observed that the estimates lie close to the initial values. This procedure has been done for all the five series and is illustrated in Table 5.1). The left column of table 5.1 represents initial values through wich data is generated and the column in the right represents the estimated values of parameters.

Table 1 Simulation study

Initial Values	**Estimates**
Model 1	
N= 300 , $\rho = 0.55,\ \alpha = 2, \sigma_1 = 3, \sigma_2 = 15, \pi_1 = 0.6,\ \pi_2 = 0.40$	$\hat{\rho} = 0.576,\ \hat{\alpha} = 1.878, \hat{\sigma_1} = 3.11,$ $\widehat{\sigma_2} = 15.527, \widehat{\pi_1} = 0.578, \widehat{\pi_2} = 0.422$
N= 500 , $\rho = 0.75,\ \alpha = 5, \sigma_1 = 3, \sigma_2 = 10, \pi_1 = 0.66,\ \pi_2 = 0.34$	$\hat{\rho} = 0.746, \hat{\alpha} = 5.073, \hat{\sigma_1} = 10.466,$ $\widehat{\sigma_2} = 15.527, \widehat{\pi_1} = 0.669, \widehat{\pi_2} = 0.331$
N= 1000 , $\rho = 0.6,\ \alpha = 4, \sigma_1 = 5, \sigma_2 = 15, \pi_1 = 0.30,\ \pi_2 = 0.70$	$\hat{\rho} = 0.583$, $\hat{\alpha} = 4.516$, $\hat{\sigma_1} = 5.011$ $\widehat{\sigma_2} = 15.083$, $\widehat{\pi_1} = 0.284$, $\widehat{\pi_2} = 0.716$
Model 2	
N= 500 , $\rho = 0.75,\ \alpha = 5, \sigma_1 = 3, \sigma_2 = 10, \sigma_3 = 4,\ \pi_1 = 0.50,\ \pi_2 = 0.30, \pi_3 = 0.2$	$\hat{\rho} = 0.583$, $\hat{\alpha} = 4.99$, $\hat{\sigma_1} = 3.012$ $\widehat{\sigma_2} = 10.671$, $\widehat{\sigma_3} = 3.753, \widehat{\pi_1} = 0.51$ $\widehat{\pi_2} = 0.296, \widehat{\pi_3} = 0.194$

N= 500 , $\rho = 0.5,\ \alpha = 3\ , \sigma_1 = 1, \sigma_2 = 7, \sigma_3 = 15,\ \pi_1 = 0.30,\ \pi_2 = 0.50, \pi_3 = 0.2$	$\hat{\rho} = 0.502\ ,\quad \hat{\alpha} = 2.99\ ,\quad \widehat{\sigma_1} = 0.931$ $\widehat{\sigma_2} = 6.528\ , \widehat{\sigma_3} = 15.577, \widehat{\pi_1}\ = 0.283$ $, \widehat{\pi_2} = 0.509, \widehat{\pi_3} = 0.208$

4. Empirical Analysis

This section presents an Empirical analysis conducted on a real-time data set. The empirical analysis justifies the applicability of our proposed model. It has been conducted to demonstrate the comparative performance of the model for different components in the mixture density and also to justify the fact that using a mixture density for innovations can be beneficial rather than using a single distribution. The dataset is sourced from Meena and Kumar (2024) (refered as Market Survey Report data set), and focuses on variables named "Sales" (y) and "Discount" (X). The variable "Sales" represents a time series with 250 data points, representing the sales of electronic items after applying discounts. The variable "Discount", as the explanatory variable, refers to the total discount offered on sales, which includes both regular discounts and negotiated bargains. the duration of the data set is September 25, 2023 to May 31, 2024. Figure (1), dipicts the time series plot of the data.

The performance of configurations with m=1,2 and 3 have been analysed. Table 1 displays the performance of various models with different values of m and Table 2 presents the parameter estimates of the models. It can be observed that using m = 2 (AIC = 1550.410, BIC = 1568.017) or m = 3 (AIC = 1554.912, BIC = 1579.563) for the model, shows improved fit as compared to m = 1 (AIC = 1656.006, BIC = 1666.570) , [Akaikie (1974)]. This indicates that employing a mixture density for innovations is a better choice than using a single density model. Particulary, the model with two Gaussian components (m=2) provides the best fit.

The Model with m=2 is the proposed model for the data under study,. (For further understanding of the structure of the proposed model it is worth mentioning that the simulated model presented in section 3 , *(model with 2-K Gaussian Innovations)* is also based on same structure.) Table (2) (second row) illustrates the parameter estimates of the model. For this model the first Gaussian component has parameter estimates $\widehat{\pi_1}$= 0.774, $\widehat{\sigma_1}$ = 3.074, the second one has $\widehat{\pi_2}$ = 0.226, $\widehat{\sigma_2}$ = 12.611. The interpretation of the parameter estimates for the proposed model is as follows: when an innovations occurs there is 77.4% probability that it came from distribution N(0,9.5) and 22.6% probability that it came from N(0,159.037). Figure (2) displays how well the proposed model (red) aligns with the original series (black). Figure (3) showcase the P-P Plot and Q-Q Plot respectively for Innovations of this model. Their alignment along a straight line justifies the fit of the mixture density.

Figure 1 Time series plot of the Sales data

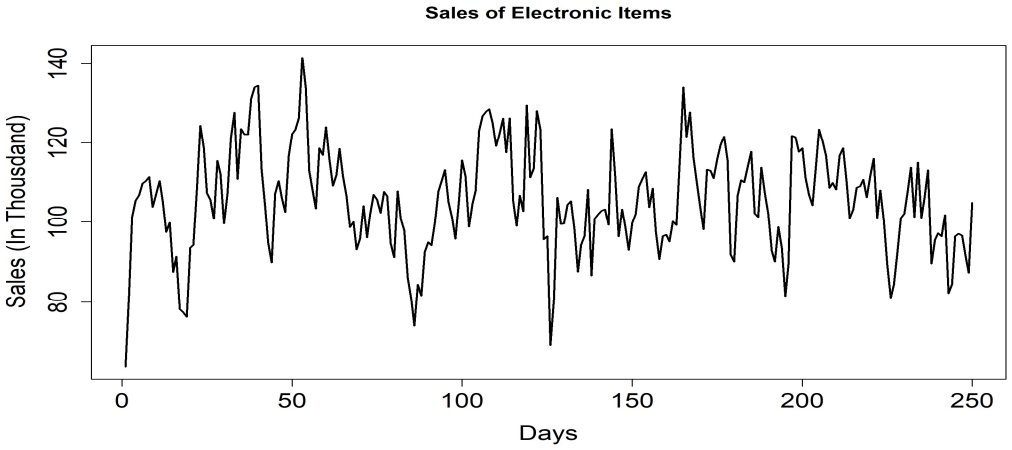


Table 1: AR (1) Models for different values of K

m (Number of Components)	AIC	BIC
1	1656.006	1666.570
2	1550.410	1568.017
3	1554.912	1579.563

Table 2 Parameter estimates for various configurations of the model

Components	Parameters
1	$\rho = 0.738, \alpha = 7.024\ , \sigma = 6.561$
2	$\rho = 0.75\ , \alpha = 6.693, \sigma_1 = 3.074, \sigma_2 = 12.611$ $\pi_1 = 0.774, \pi_2 = 0.226$
3	$\rho = 0.75\ , \alpha = 6.614\ \sigma_1 = 3.059, \sigma_2 = 12.645$ $\sigma_3 = 3.111, \pi_1 = 0.516, \pi_2 = 0.225,$ $\pi_3 = 0.259$

Figure 2: Original Time Series and proposed Model Comparison

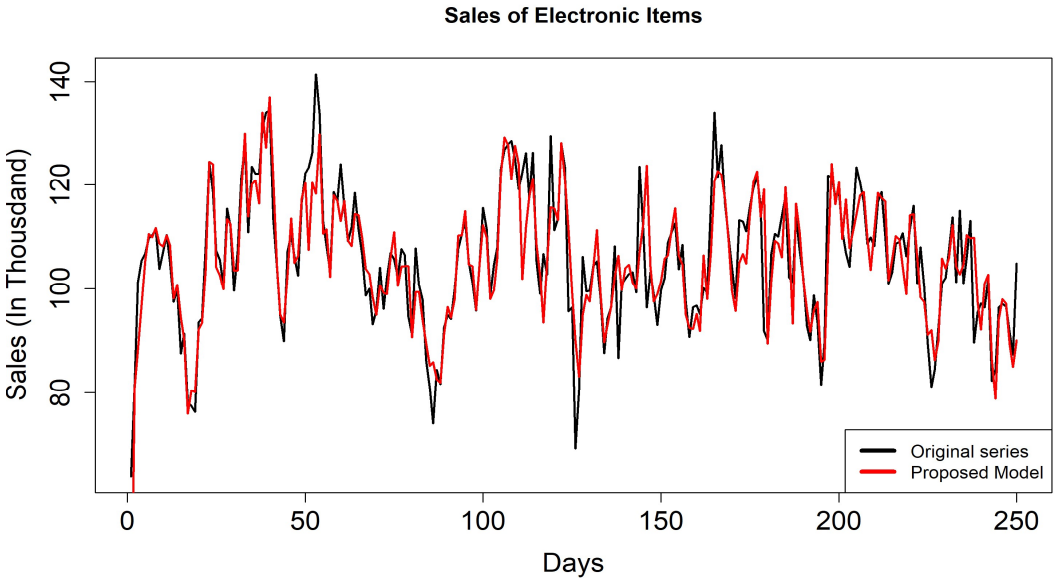
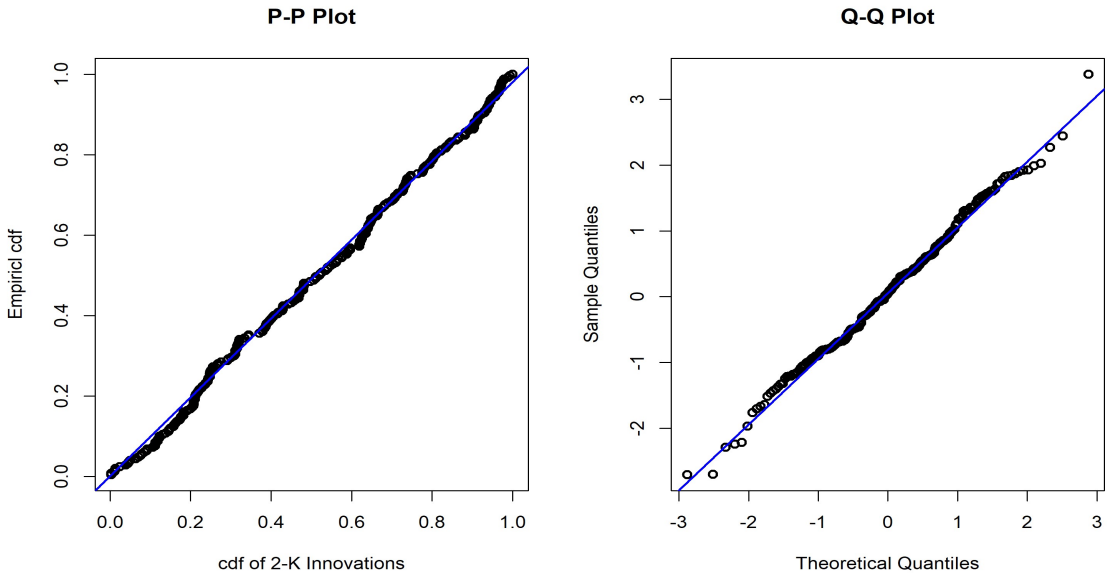


Figure 3: P-P Plot and Q-Q Plot for the Innovations from the proposed model



5 Conclusion

This study has explored the AR(1) model with an explanatory variable and innovations governed by a finite mixture of Gaussian distributions. The simulation study demonstrated that this advanced model configuration offers significant improvements in capturing the underlying dynamics of time series data. The mixture-distributed innovations allowed for a more flexible and realistic representation of the error structure, leading to enhanced predictive accuracy and robustness.

The findings suggest that incorporating mixture distributions in the innovations can effectively address issues such as clustering, making the model particularly useful for complex datasets encountered in various fields.

In this study , the focus was solely on the Gaussian distribution for modeling the time series. However, it is essential to recognize that the underlying theory can be extended to include other appropriate distributions as well. Different distributions might be better suited to capture specific characteristics of the data, and exploring alternative families could offer valuable insights into model performance and flexibility. Furthermore, the present study is based on the assumption of using the same distribution family for each component in the mixture model. While this simplification was useful for our initial investigation, future research could delve into the potential advantages of considering different distribution families for distinct components.

Conflict of Interest:

There is no conflict of interest regarding the publication of this paper.

Funding:

No funding was received for conducting this study.

Data Availability Statement:

No new data were generated or analysed in this study. This study is based on previously published data and materials, all of which are cited appropriately in the references section.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

An, H. Z., & Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. Statistica Sinica, 943-956.
Anderson, T. W., (1958). *An introduction to multivariate statistical analysis* (Vol. 2, pp. 3-5). New York: Wiley.
Balakrishna, N. (2022). Some Non-linear AR-type Models for Non-Gaussian Time Series. In Non-Gaussian Autoregressive-Type Time Series (pp. 127-154). Singapore: Springer Nature Singapore.
Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Brockwell, P. J., & Davis, R. A. (1991). Time series: Theory and methods. Springer-Verlag

Chandra, S. (1977). On the mixtures of probability distributions. Scandinavian Journal of Statistics, 105-112.Chatfield, C., & Xing, H. (2019). The analysis of time series: an introduction with R. Chapman and Hall/CRC.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.

Douc, R., Moulines, E., & Stoffer, D. (2014). Nonlinear time series: Theory, methods and applications with R examples. CRC Press.
Everitt, B. (1996). An introduction to finite mixture distributions. Statistical Methods in Medical Research, 5(2), 107-127.
Everitt, B. (2013). Finite mixture distributions. Springer Science & Business Media.
Everitt, B. S., & Hand, D. J. (1981). Mixtures of normal distributions. In Finite Mixture Distributions (pp. 25-57). Dordrecht: Springer Netherlands.
Fan, J., & Yao, Q. (2008). Nonlinear time series: nonparametric and parametric methods. Springer Science & Business Media.
Hamilton, J. D. (2020). Time series analysis. Princeton University Press.
Jones, D. A., & Cox, D. R. (1978). Nonlinear autoregressive processes. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 360(1700), 71-95.
Kumar, S., & Hooda, S. . (2023). Healthcare and Hospital Promotions and Audience Reception. *MediaSpace: DME Media Journal of Communication*, *3*(01), 16–19. https://doi.org/10.53361/dmejc.v3i01.03
Kumar, S., & Singh, P. (2023). Emotional Appeal in the Tweets: A Study on Indian National Political Parties. *Journal of Communication and Management*, *2*(02), 95–97. https://doi.org/10.58966/JCM2023223
McLachlan, G. J., & Krishnan, T. (2008). The EM algorithm and extensions. John Wiley & Sons.
Meena R, Kumar A. Impact of discounts on sales in Delhi NCR: a survey report. Int J Sci Rep 2024;10(11):405-9
Miljkovic, T., & Grün, B. (2016). Modeling loss data using mixtures of distributions. Insurance: Mathematics and Economics, 70, 387-396.
Pearson, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A, 185, 71-11.
Rao, T. S., Rao, S. S., & Rao, C. R. (Eds.). (2012). Time series analysis: methods and applications (Vol. 30). Elsevier.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). Time series analysis and its applications (Vol. 3, p. 4). New York: Springer.
Sudhir Kumar, & Manpreet Singh. (2021). The News Is Either Restricted Or Excessively Political: A Perception Study Of People Of Delhi. *Elementary Education Online*, *20*(1), 4528–4538. Retrieved from https://ilkogretim-online.org/index.php/pub/article/view/2047
Tjøstheim, D. (1990). Non-linear time series and Markov chains. Advances in Applied Probability, 22(3), 587-611.
Zakoian, J. M. (1992). Howell (Tong)—Nonlinear Time Series. A Dynamical System Approach. Population, 47(5), 1320-1321.