STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, February 2025, pp 741–758.
Published online in International Academic Press (www.IAPress.org)

Comparison between the FUZZY-ARFIMA Model and the Hybrid ARFIMA-FUZZY Model with Application to Agricultural Data in the City of Mosul

Rehab Talal Ahmed, Omar salim Ibrahim*

Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Iraq

Abstract In this research, we studied forecasting based on time series data for red onion prices in Nineveh Governorate using model ARFIMA Autoregressive fractionally integrated moving average. A ARFIMA-FUZZY (FTS) hybrid model was proposed This model has the advantage and strength of the ARFIMA partial autoregressive integral in addition to the FUZZY-ARFIMA model and compared them with each other using evaluation criteria (BIC). For prediction, which is calculated using the statistical program R. The results showed that the ARFIMA-FUZZY (FTS) hybrid model is the best because it has the lowest (BIC) values. It is also the highest in forecast efficiency because it has the lowest values of forecast accuracy metrics (MSE, RMSE, MAE) and was chosen as the best forecast model.

Keywords Fuzzy Time Series, Membership Function, Predictive, Hybrid Model, ARFIMA .

AMS 2010 subject classifications 62A86, 62M10

DOI: 10.19139/soic-2310-5070-2092

1. Introduction

The time series forecasting process is considered one of the most important methods of statistical analysis that plays a major role in making decisions under uncertainty. A time series is defined as a group of related historical measurements and observations of a phenomenon for certain periods of time and are usually equal in length. In recent decades, researchers have been interested in building time series models due to their importance and their ability to explain phenomena and data in many fields, including (economic, social, medical, etc.) and even at the individual level. The most important of these models are the long-memory time series model and the FTS model. The FTS model is characterized by a flexible mechanism that simulates... Sometimes basic time series models are used because they do not require the basic assumptions to build the model in the forecasting process. This model is characterized by high prediction accuracy, In practice, most time series are characterized by two components, linear and nonlinear, and when making predictions, individual models are not sufficient to model these series. Recently, several linear, nonlinear, and hybrid models have been proposed for forecasting, in this regard In the research, a new hybrid model was proposed based on combining the linear regression partial integral moving average (ARFIMA) model with the nonlinear fuzzy-time model. String model (FTS). The proposed hybrid model analyzes the linear component of fixed time Using the ARFIMA model, you calculate the estimated values, and then calculate the residual values For this model by subtracting the estimated values from the original time series. Non-linear The component is analyzed using the calculated residual (FTS) model, which is inherently contained Nonlinear patterns of time series. Final values for prediction through hybrid application The ARFIMA-FTS model

*Correspondence to: Omar Salim Ibrahim (Email: omarsalim85@uomosul.edu.iq). Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

is obtained by combining the predictions of the ARFIMA model The original series with FTS model predictions of the remaining series. New hybrid.

Many studies have addressed this type of model and its application in many different fields, including study [6] ”The hybrid model of autoregressive fractionally integrated moving average and fuzzy time series Markov chain on long-memory data” The research aims to determine the long pattern of crude oil price movement through a partial time series model where the accuracy can still be improved by constructing a hybrid residual model using a fuzzy time series approach. Long memory data indicate a high level of volatility and autosequence value between lags that Decrease slowly. However, a more accurate model is proposed as a hybrid time series model with fuzzy time series Markov chain (FTSMCA) time series model of crude oil price is obtained as a new target model for the hybrid model of ARIMA and ARFIMA with FTSMC, known as ARIMA-FTSMC and ARFIMA-FTSMC, respectively. The exact model measured by MAE, RMSE and MAPE shows that the hybrid model of ARIMA -FTSMC has better performance than ARIMA and ARFIMA, but the hybrid model of ARFIMA-FTSMC provides the best accuracy compared to all models. The superiority of the hybrid temporal model of ARFIMA-FTSMC on long-memory data provides an opportunity for the hybrid model as the best and most accurate prediction method. Also study [15], ”New hybrid fuzzy time series model: Forecasting the foreign exchange market” this work compares the forecasting of volatility for traditional time series models (ARIMA, EGARCH, and PARCH) against two proposed new models based on fuzzy theory (FTS Fuzzy ARIMA Tseng’s and FTS-Fuzzy ARIMA Tanaka’s). The aim of the study is that models based on fuzzy theory generate better estimates of volatility. Fuzzy models show lower prediction errors compared to traditional time series in both in-sample and out-of-sample. The models demonstrate higher efficiency and better reflect market information.

2. Search goal

It aims to compare between ARFIMA models, FUZZY ARFIMA and ARFIMA FUZZY, choose the best model, thrrough predict the prices of red onions in the city of Mosul.

3. Long memory process

The basic models that allow determining long memory are the Autoregressive fractionally Integrated Moving Average (ARFIMA) [9, 8] and are considered an extension of the (ARFIMA) models that take the fractional differential coefficient d as real values for the period (-0.5,0.5). The mathematical formula of the ARFIMA model can be expressed using the (Wold) relationship as follows:

$$\gamma_t = \sum_{j=0}^{\infty} \psi j \ \varepsilon t_j \tag{1}$$

γ_t: Time series.
ψj: Moving average weights $\psi j \epsilon R, \psi \ 0 = 1$.
εt: It is the process of white noise, as follows: White Noise, where:
Var(εt)= σ_ε^2, $E(\varepsilon t) = 0$, εt i.i.d $(0, \sigma_\varepsilon^2)$
A time series is considered to have a long and static memory if:

$$\sum_{j=0}^{\infty} \left| \psi^2 \ j \ \right| = \infty \tag{2}$$

Therefore, it can be said that any time series is an ARFIMA model (p, d, q) if the following condition is met:

$$\Phi_P \left(L \right) \left(1 - L \right)^d y_t = \Theta_q \left(L \right) \tag{3}$$

$$\Phi_P(L) = 1 - \sum_{i=1}^{P} \Phi_i L^i = 1 - \Phi_1 L - \cdots \Phi_P L^P$$
$$\Theta_q(L) = 1 - \sum_{j=1}^{q} \Theta_j L^j = 1 - \Theta_1 L - \cdots \Theta_q L^q \tag{4}$$

$(\Theta\ L)$ and $(\Phi\ L)$ respectively represent polynomials in the two parts MA(q) and AR(P) of the model p, q respectively. $(1-L)^d$ the back displacement factor, d the fractional differential coefficient. Also, the properties of ARFIMA models can be clarified according to the different values of the fractional differential coefficient (d) when $d < 1/2$ and the roots of the polynomial $\Theta_q(L)$ lie outside the unit roots, in this case the series y_t is invertible. When $d < -1/2$, all the roots of the characteristic polynomial $\Theta_q(L)$ lie outside the unit root, in this case the series is sTable, and when $-1/2 < d < 0$ the series y_t is reversible and has an incomplete short memory. While $0 < d < 1/2$, the series y_t is a static series with long memory (long-term stability), and it is also continuous, as the positive autocorrelation function slowly decreases towards zero "in the form of a hyperbola" as the number of gaps k increases.

3.1. *Testing of long memory*

In our research, many statistical tests were used to verify long memory, including the Geweke_Porter_Hudak estimator, GPH, the Smoothed_priodogram_estimator, (dsprio), the Fracdiff method, the Rescaled_Range, R/S, and the Whittle_estimator. The model was also built by verifying the presence of long memory in the time series through several tests and then moving on to estimating the fractional difference parameters. We will explain the (Rescaled_Range, R/S) method [3].

[11] used a statistic to examine whether a time series has a long memory (long-term correlations), as the goal of the R/S statistic is to calculate the Hurst coefficient, and the R/S statistic is defined as the range R of the partial sums of deviations for a series. time from the mean divided by its standard deviation ST.

$$R/S = Qn = \frac{1}{\sigma_y}\left[\max_{1\le k\le n}\sum_{j=1}^{k}(yj - \bar{y}n) - \min_{1\le k\le n}\sum_{j=1}^{k}(yj - \bar{y}n)\right] \tag{5}$$

k: The number of partial sums between the items of the series y_t and its arithmetic mean $\bar{y}_t$. R/S analysis allows calculating what is called the Horst coefficient $1H << 0$, which is defined as the ratio between the logarithm of R/S and the logarithm of the number of observations T:

$$H \approx \frac{\log Qt}{\log T} \tag{6}$$

[9] and [13] showed that there is a strong relationship between the fractional differential coefficient (d) of ARFIMA [12] models and the Horst coefficient (H), as:

$$d = H - \frac{1}{2} \tag{7}$$

Through the previous relationship, it is possible to determine the fractional differential parameter d and determine whether the string has a long memory, and this depends on the variable values of the factor H, as follows:

1. When $H = 1/2, d = 0$. In this case, there is no connection between the phenomena of past and present events, that is, it is an ARMA model.
2. When $1/2 < H < 1$ is $0 < d < 1/2$, then the ARFIMA model is considered long-memory stable as the correlations are stronger the closer H is to one.
3. When $0 < H < 1/2$ is $(-1/2) < d < 0$, in this case the chain has a long memory and at the same time does not behave like ARMA models.

3.2. Estimation of ARFIMA model

There are two methods for estimating ARFIMA models. These methods are divided into two-stage methods, where the fractional differential parameter (d) is estimated and then the ARMA parameters are estimated. These methods include (Geweke_Porter_Hudak estimator, (Smoothed priodogram estimation ”dsprio”, Fracdiff), and the method One-stage estimation, where the fractional differential parameter is estimated simultaneously with the estimation of the ARMA parameters. Single-stage estimation methods are considered the most effective methods. This method requires that the series be stable or converted to stable. Among these methods is (EML) Exact Maximum Likelihood, which is considered one of the most effective methods for estimating a parameter. The fractional difference (d) in parallel with the AR, MA parameters of the ARFIMA model. The condition of this model is that the series be stable or be converted to stable. This method allows the use of all long-term and short-term information associated with the time series.

3.3. Diagnostic

This stage is considered one of the most important stages followed in developing the model to provide the possibility of applying model information for prediction by matching its parameters with statistical hypotheses, so that diagnosing the model in general depends on conducting many tests including testing and analyzing its stability by examining estimates of correlation coefficients. Subjectivity obtained from the estimation stage.

3.4. Forecasting

Forecasting is the last stage of time series analysis and can only be reached after tests to diagnose the model. After obtaining a suiTable model to represent the data, the model becomes ready to be used to predict future values [19]. Criteria or (evaluation metrics) are used in this stage. We achieved the accuracy of the model and its ability to produce efficient data. Below are some of these criteria(AIC,BIC,MSE,RMSE,MAE).

4. Fuzzy time series concepts and models

Fuzzy logic serves as the foundation for the FTS approach as it primarily relies on fuzzy sets used by the algorithms of these models to operate. Below is a summary of the definitions of FTS as in [18].

1. Fuzzy Time Series
   Let $X(t)(t = 0, 1, 2, \ldots.)$ be one of the subgroups of real numbers, with fuzzy sets $Ai(t)(t = 0, 1, 2, \ldots.)$ subsequently defined on them. The fuzzy time series $X(t)$ is then defined as $F(t)$, which is a collection of $Ai(t)$.
2. Fuzzy Relationship
   Let $F(t)$ be a fuzzy time series if $F(t)$ is the result of $F(t-1)$, then $F(t-1) \to F(t)$ represents this fuzzy logical relationship and is known as a first-order fuzzy time series model.
3. Fuzzy Relationship by Order N
   Let $F(t)$ be a fuzzy time series, and if $F(t)$ is the result of $F(t-1), F(t-2), \ldots, F(t-N)$, then this fuzzy logical relationship is expressed as follows:
   $F(t-N), \ldots, F(t-2), F(t-1) \to F(t)$

4.1. Fuzzy logic sets

In fuzzy logic, sets can be of two types:

4.1.1. Actual crisp set According to [12], this is a set of elements with distinct characteristics that may or may not be defined, may or may not belong. To distinguish them from fuzzy sets in terms of definition, these elements are called actual or traditional sets and have two values: (1) when they belong to the set or (0) when they do not.

4.1.2. Fuzzy set Defined as a set consisting of multiple types (categories) of members with a membership function ranging between 0 and 1 [20]. The value zero indicates that the element is not part of the set or that the degrees range between 0 and 1.

4.1.3. Membership function Fuzzy logic uses membership functions that distinguish one from another by describing whether it is continuous or discrete [16]. The range of values is between (0,1), where 0 means the value does not belong to the set, 1 means the value belongs to the set, and among the membership functions we mention the following:

1. Triangular membership function
   The mathematical function can be expressed through the following formula:

$$\mu_A(x) = \left\{ \begin{array}{ll} 1 - \frac{|x-a|}{c}, & a-c \leq x \leq a+c \\ 0, & \text{otherwise} \end{array} \right\} \tag{8}$$

2. Sigmf (sigmoidal membership function)
   This function is open from the right to represent the largest wave value and from the left to represent the largest negative value.
3. Bell membership function
   The mathematical formula for this function is as follows:

$$\mu_A(x) = e^{-\frac{(x-a)^2}{2b^2}} \tag{9}$$

4. Trapezoidal membership function
   The mathematical formula for this function is as follows:

$$\mu_A(x) = \left\{ \begin{array}{ll} \frac{(a-x)}{(a-b)}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{(d-x)}{(d-c)}, & c \leq x \leq d \\ 0, & \text{otherwise} \end{array} \right\} \tag{10}$$

4.2. *Fuzzy time series model*

The initial stage of creating an FTS model involves dividing the universe of study (U), described by the symbol (X) into a series of fuzzy sets (periods), where $U = [\min(x), \max(x)]$ Additionally create the linguistic variable A, composed of the fuzzy sets $Aj, j = 1, 2, .., k$, [4].

In the world of study U, a fuzzy set can be expressed as:

$$A = fA(u1)/u1 + fA(u2)/u2 + \ldots + fA(un)/un(4)$$

Where fA denotes the membership function of the fuzzy set A, $fA : U \to [0, 1]$ and $fA(ui), (1 \leq i \leq n)$ denotes the degree to which ui is a member of $fA(ui)\epsilon[0, 1]$.

The traditional way that has been presented in many studies to end this stage is to use fuzzy organic functions, as it has appeared in many studies [7], U is divided into equal fuzzy sets, and the three most important membership functions in this process are Gaussian, trapezoidal, and trigonometric membership functions.

4.3. *Predictive steps using FTS model*

The key steps identified in the stages of the prediction process, according to the FTS model, are as follows [18]:

1. Description and division of the study domain.
2. Fuzzification process.
3. Identification of fuzzy relationships.

4. Defuzzification method.

Below is a diagram illustrating the system for inputs (Xt) being fuzzified, i.e., converted into membership grades (MF) in the fuzzy input set. It then moves to the fuzzy inference engine using IF, THEN rules stored in the rule base, the fuzzy inference engine produces fuzzified values, resulting in a usable outcome.

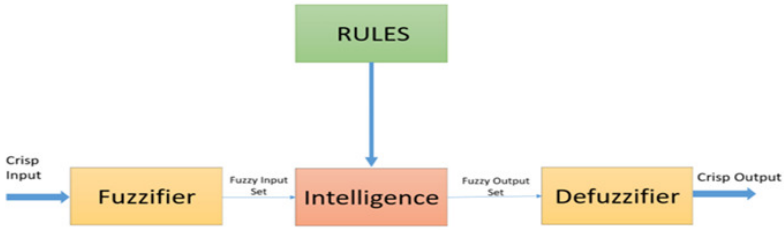

Figure 1. Fuzzy logic diagram.

4.4. Main FTS models

The goal of this section is to provide a brief overview of common fuzzy time series models. Models introduced by [18, 4, 10], and [1], the model [5] and the model [17].

5. Hybrid (ARFIMA-FTS) model

It is known that time series are of two types: linear and non-linear time series. Linear time series assume a linear data generation process such as ARIMA and ARFIMA. Despite the flexibility of these models, they cannot deal with non-linear data. Assuming linearity in practical reality is something that cannot be accepted. In all cases [14], and quite the opposite, building non-linear time series models is very suiTable for practical reality. Given the difficulty of modeling these patterns, a method was proposed by combining linear and non-linear components, which was called hybrid forecasting models, which can be It is effective for improving predictions, and since time series often combine several problems, this prompted researchers to combine and merge models together in order to solve all problems simultaneously using one comprehensive model, as the hybrid model maintains the accuracy of predictions in addition to preserving the trend of the data [2].

A hybrid method was used and a model for a long memory time series was proposed by (G.P. Zhang) through the ARFIMA_FTS approach, which considered the time series as consisting of two patterns, the linear pattern and the non-linear pattern, and the formula for this model is as follows:

$$Y_t = L_t + N_t \tag{11}$$

whereas: Y_t: represents the time series.
L_t: Autoregressive linear composite of the time series.
N_t: The nonlinear random error component of the time series.
The prediction steps of the Mahdist model can be summarized as follows:

- Estimating the best linear model ARFIMA for the time series
- Calculating the residuals for the ARFIMA model is calculated according to the following formula:

$$e_t = y_t - \hat{L}_t \tag{12}$$

Where e_t represents the residuals of the ARFIMA model at time t.

$$e_{t=}f\left(e_{t-1}, e_{t-2}, \ldots, e_{t-n}\right) + \varepsilon_t \tag{13}$$

Where f is a non-linear function estimated using the FTS model and ε_t is the random error. By neglecting it, we obtain predictions and they are calculated according to the following formula:

$$\hat{N}_{t=}f\left(e_{t-1}, e_{t-2}, \ldots\ldots, e_{t-n}\right) \tag{14}$$

The predictive values of the time series are calculated according to Equation (4) The hybrid model has the advantage of combining the linear-style ARFIMA model with the FTS model. The results of the hybrid model are often more satisfactory, especially in the long run, than using the ARFIMA and FTS models alone. Despite these advantages, it cannot be said with certainty that this model is better than the ARFIMA model, especially in Short-term.

6. The applied aspect

6.1. Applying ARFIMA model

In this study, the prices of the red onion crop in Iraqi dinars will be predicted in Nineveh Governorate using original data for the years from January 2018 to December 2023, on a weekly basis, which were obtained from the Nineveh Agriculture Department, The R language was used to write the program for the model building phase used in this thesis to reach the final results, Using packages (tseries, pracma, fracdiff, longmemo, AnalyzeTs, fuzzy.ts1, fuzzy.ts2), Table 1 shows the most important descriptive statistics for the time series.

Table 1. Descriptive statistics for the time series

statistics	Sample size	mean	Median	Variance	Minimum value	highest value
Valuable	288	480	492	53415.17	200	1500

6.2. Stages of the ARFIMA methodology

6.2.1. Chain stability check We will draw the time series of imported Red onions crop price data and draw the autocorrelation function and the partial autocorrelation function. This step aims to examine the stability of the time series.

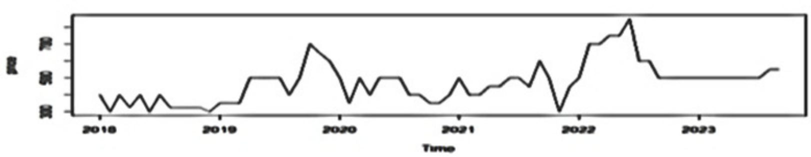

Figure 2. Time series of Red onions price.

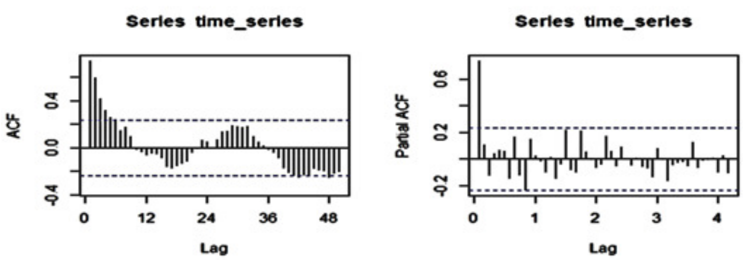

Figure 3. Autocorrelation and partial autocorrelation for Red onions price.

We notice by drawing both ACF and PACF that the series is unstable, and it is known that there are tests to examine the stability of the time series, and this will be done through the expanded Dickey-Fuller (ADF) test, and the Philip-Byrne (PP) test. The Table shows these tests. The statistical significance of the P-value for the ADF test

Table 2. Is a stability test for time series

Statistical test	value	P-value	Status of the series
ADF	-2.8911	0.2127	Un Stable
PP	-26.269	0.1047	Un Stable

reached (0.2127), which is greater than 0.05, and also for the Philip_Bern test reached (0.1047), meaning that the P-value is > 0.05, and this means that the time series is unstable.

6.2.2. Long-term memory tests It was noted from the Table that if the value of 1/2 $< H < 1$, the time series has a long memory and that the correlations are strong the closer H is to one.

Table 3. Long-term memory test

Test	R/S	Empirical Hurst exponent	Theoretical Hurst exponent	Whittle
H	0.7002161	0.6948797	0.5179464	0.9494125

6.2.3. Methods for estimating the fractional difference factor (d) The following is the estimation of the fractional differential coefficient by direct and indirect methods, which depend on the Horst coefficient H, and the value of d is calculated in the R/S, Whittle test according to the formula $d = H - 1/2$, and in direct methods, d is calculated directly such as the dSperio, Gph, Fracdiff test.

Table 4. Methods for estimating (d)

Test	R/S	Empirical Hurst exponent	Theoretical Hurst exponent	Gph	dSperio	Fracdiff	Whittle
value (d)	0.2002161	0.1948797	0.0179464	0.5077225	0.4663449	0.472136	0.4494125

From the table above it is clear that all test results fall within the interval limits (0, 0.5), meaning that the time series is characterized by long memory.

6.2.4. Recognition stage This step is important and is the first stage of time series analysis, through which we learn about the model, that is, choosing the ranks of the model (p, d, q). Since the series is unstable, we will use the fractional differential coefficient d that we obtained from the previously mentioned methods. We test the stability of the series after taking the five fractional differences that passed the long memory test using the PP test because it is better and more accurate than the ADF test.

Table 5. Results of PP test for the series after taking the fractional differences of the previous methods

Estimation methods	value	P-value	The situation
R/S	-37.334	0.01	stable
Empirical Hurst exponent	-36.998	0.01	stable
Theoretical Hurst exponent	-27.371	0.01	stable
Fracdiff	-58.593	0.01	stable
Whittle	-56.141	0.01	stable

The results of the pp tests shown in Table 5 prove that the time series is stable and can be used in building ARFIMA models, as the moral value of the five methods test was smaller than 0.05.

6.2.5. Diagnosis and assessment The goal of this step is to identify one or more ARFIMA models by knowing the rank of AR(P) and rank of MA(q), where we will use the (BIC) criterion to compare the models and to estimate the methods used for fractional differences.

In our study, we will use the estimation of ARFIMA models using the two-stage and one-stage estimation method and compare the models to obtain the best prediction model that has the lowest value of the BIC criterion and the prediction accuracy criteria MSE, RMSE, MAE, as shown in the Table below.

Table 6. Comparison of ARFIMA models with two-stage and one-stage method

Model	BIC	Φ1	Φ2	Θ1	Θ2	MSE	RMSE	MAE
(1, 0.2002161, 0)	645.1705	0.8699	0	0	0	6227.978	78.91754	58.22019
(1, 0.1948797, 0)	645.1625	0.8760	0	0	0	6226.117	78.90574	58.22019
(1, 0.0179464, 0)	643.311	0.885336	0	0	0	6137.466	78.30578	58.15908
(1, 0.5077225, 0)	644.7165	0.81587	0	0	0	6295.565	79.3446	59.4843
(1, 0.4663449, 0)	644.7074	0.8159	0	0	0	6287.602	79.2944	59.48312
(1, 0.472136, 0)	644.827	0.82371	0	0	0	6288.66	79.30107	59.48709
(1, 0.4494125, 0)	644.8385	0.8246	0	0	0	6284.577	79.27533	59.4645
(1, 0.0000458, 0)	645.3388	0.885333	0	0	0	6148.706	78.34989	58.16395

By comparing the two-stage long memory estimation models with the one-stage long memory estimation model, it turns out that the best model is (1, 0.0179464, 0) with the two-stage estimation method, as it has the lowest BIC standard, as well as the lowest value for MSE, RMSE, MAB.

Table 7. Estimate Model Parameters (1, 0.0179464, 0)

model	Parameters	Appreciation	p.value
AR(1)	Φ1	0.885336	0.0000

The mathematical formula for the model (1, 0.0179464, 0) ARFIMA is as follows:

1. Residual Analysis ARFIMA (1,0.0179464,0)
   Residual analysis is considered the most important stage to determine the suitability of the model (1, 0.0179464, 0) ARFIMA for use in forecasting.

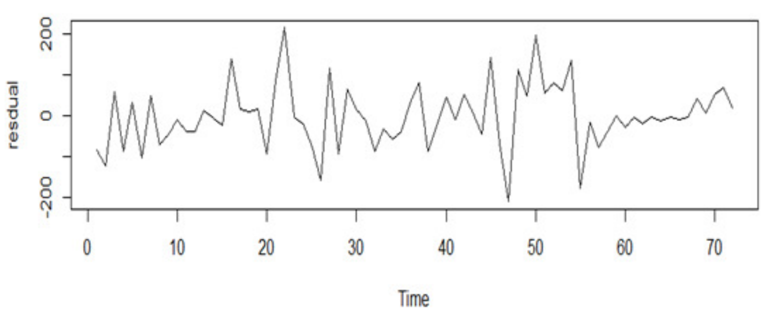

Figure 4. Residual curve for the data of the studied series.

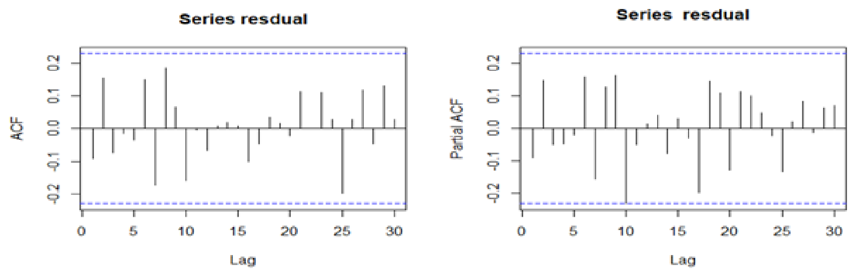

Figure 5. The autocorrelation function of the residuals and the partial autocorrelation function of the residuals.

2. Ljung box test
   Below is a Table showing the test results for the residuals of the model (1,0.0179464,0):

Table 8. Ljung box test

Model	Q*	p-value
(1,0.0179464,0)	12.729	0.548

We note that the significant value is ¿ 0.05 p-value, which means that the residuals are independent and that the model has passed the diagnosis.

6.2.6. Prediction: Prediction for the ARFIMA(1, 0.0179464, 0) model After choosing ARFIMA (1, 0.0179464, 0) as the best model in diagnosis and the best representative model for the data, we used it for prediction. The prediction will be for 12 values and then compare them with the real values, as shown in the table below.

Table 9. Ljung box test

Sequence	1	2	3	4	5	6	7	8	9	10	11	12
True value	350	450	400	500	400	500	500	400	400	400	500	500
Predicted value	578	554	536	522	512	504	498	494	490	488	486	578

6.3. Applying FTS or FUZZY- ARFIMA model

After verifying the stability of the time series for the price of red onions and conducting long memory tests, it was confirmed that the series has a long memory, and the best model was determined by conducting a comparison between the models using the comparison standard BIC and MSE, RMSE, MAE, where the results showed that the model ARFIMA (1, 0.0179464,0) has the lowest Criteria values, we will now fuzzify the string data using fuzzy analysis. singh's time series fuzzing algorithm was used to address the fuzziness of the local red onion price in the city of Mosul and predict it for the coming period. This was done through a series of the following steps:

1. Determine the lower and upper limit of the red onion price data series, where the lower limit was (200) and the upper limit was (1500), and by applying the equation, $U = D_{\min} - D_1, D_{\max} + D_2$, where $D_{\min} = 200\ and D_{\max} = 1500, D_1 = 0, D_2 = 0$ then $u = [200, 1500]$.
2. Partitioning the red onion price data into equal length periods $(u_1, u_2, \ldots, u_{13})$ followed by the labeling of fuzzy sets denoted by $(A_1, A_2, \ldots, A_{13})$ as shown in the Table below.

Table 10. The periods and the fuzzy sets

Periods	Periods values	Fuzzy set
U_1	[200,300]	A_1
U_2	[300,400]	A_2
U_3	[400,500]	A_3
U_4	[500,600]	A_4
U_5	[600,700]	A_5
U_6	[700,800]	A_6
U_7	[800,900]	A_7
U_8	[900,1000]	A_8
U_9	[1000,1100]	A_9
U_{10}	[1100,1200]	A_{10}
U_{11}	[1200,1300]	A_{11}
U_{12}	[1300,1400]	A_{12}
U_{13}	[1400,1500]	A_{13}

3. Fuzzing the data, as each value will be fuzzy in the fuzzy group that belongs to it.
4. Calculating the sums of fuzzy relationships when a set of several fuzzy sets are connected with each other by a certain set, the right side of them is merged to form the Fuzzy Relationship Groups, the Table below illustrates the fuzzy relationships:
5. Processing the expected output fuzziness is done in two steps, the first step is calculating the midpoints of the time periods U, and in the second step, processing the fuzziness using the averaging method and by applying the defuzzification rules to the time series and extracting the prediction values for the red onion price series. The values were obtained using the R program.
   After obtaining the fuzzy values, the FUZZY-ARFIMA methodology is applied:

Table 11. Fuzzy relationship sets

Groups	The fuzzy relationship is second-order
G1	A1→A1,A2,A3,A4
G2	A2→A1,A2,A3,A4
G3	A3→A1,A2,A3,A4,A5,A6
G4	A4→A1,A2,A3,A4,A5,A6
G5	A5→A3,A4,A5,A6,A10
G6	A6→A3,A4,A5,A6,A7,A8
G7	A7→A4
G8	A8→A6,A8,A10,A12,A13
G9	A9→NA
G10	A10→A5,A8,A11
G11	A11→A8,A10,A13
G12	A12→A10,A11
G13	A13→A8,A11,A13

6. Processing the expected output fuzziness is done in two steps, the first step is calculating the midpoints of the time periods U, and in the second step, processing the fuzziness using the averaging method and by applying the defuzzification rules to the time series and extracting the prediction values for the red onion price series. The values were obtained using the R program.

After obtaining the fuzzy values, the FUZZY-ARFIMA methodology is applied:

6.3.1. Stability test The figure below shows a plot of the time series:

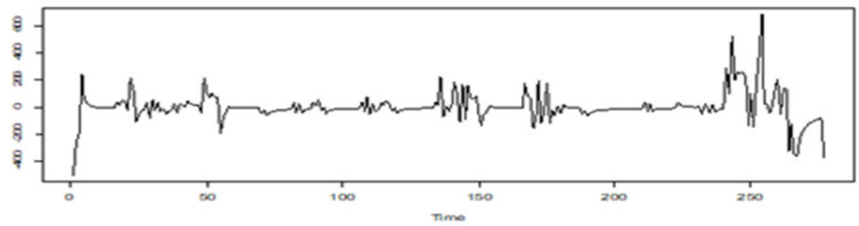

Figure 6. Fuzzy time series.

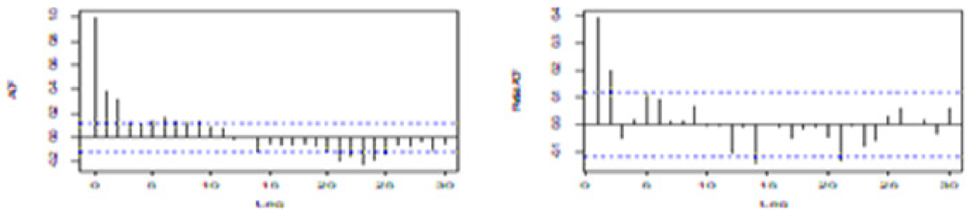

Figure 7. Autocorrelation and partial autocorrelation of the fuzzy series.

To ensure we check stability by conducting tests:

Table 12. Is a stability test for fuzzy series

Statistical test	value	P-value	Status of the series
ADF .test	-2.5684	0.01	stable stable
PP .test	-28.645	0.01	stable

It is clear from the data of the Table above that the statistical significance of the P-value for the ADF. test reached (0.01) which is less than 0.05, and P-value for the pp .test is (0.01).

6.3.2. Long-term memory tests It was noted from the Table that the value of H= 0.7552436 in the R/S test, and since it is known that if $1 < H < 1/2$, the time series is characterized by a long memory and that the correlations are strong the closer H is to one, as well as in the Theoretical Hurst exponent, Empirical Hurst test. exponent, Whittle Therefore, the string has a long memory feature.

Table 13. Long-term memory tests for fuzzy time series of red onion price by Singh method

Test	R/S	Empirical Hurs texponent	Theoretical Hurst exponent	Whittle
H	0.7552436	1.073825	0.5538539	0.9899553

It was noted from the Table that the value of H= 0.7552436 in the R/S test, and since it is known that if $1 < H < 1/2$, the time series is characterized by a long memory and that the correlations are strong the closer H is to one, as well as in the Theoretical Hurst exponent, Empirical Hurst test. exponent, Whittle Therefore, the string has a long memory feature.

6.3.3. Methods for estimating the fractional difference factor (d) The following is an estimation of the fractional differential coefficient using direct and indirect methods that depend on a coefficient Hurst H and the value of d is calculated in the R/S, Whittle test according to the formula $d = H - 1/2$, and in Direct methods d are calculated directly such as dSperio, Gph, Fracdiff test.

Table 14. Methods for estimating (d)

Test	R/S	Empirical Hurst exponent	Theoretical Hurst exponent	Gph	dSperio	Fracdiff	Whittle
value (d)	0.2552436	0.573825	0.0538539	0.6918797	0.8308542	0.4972547	0.49

The Gph, and dSperio method, are excluded because the value of d falls outside the range (0, 0.5), while the remaining methods fall within the range (0, 0.5), indicating long-range memory in the time series.

6.3.4. Stage of recognition We will conduct a(PP) test to examine the stability after taking the fractional differences:

Table 15. Results of PP test for the series after taking the fractional differences of the previous methods

Estimation methods	value	P-value	The situation
R/S	-92.861	0.01	Stable
Empirical Hurst exponent	-212.34	0.01	Stable
Theoretical Hurst exponent	-37.125	0.01	Stable
Fracdiff	-187.32	0.01	Stable
Whittle	-184.71	0.01	Stable

The results of the pp. tests shown in Table 15 prove that the time series is sTable in PP and can be used in building ARFIMA models for fuzzy values, as the significance value of the PP. test for the above methods was smaller than 0.05.

6.3.5. Diagnosis and assessment The goal of this step is to identify one or more ARFIMA models by knowing the AR(P) rank and MA(q) rank, where we will use the (BIC) criterion to compare the models and to estimate the methods used for fractional differences. compare the models to obtain the best prediction model that has the lowest value of the BIC criterion and the prediction accuracy criteria MSE, RMSE, MAE, as shown in the Table below

Table 16. Results of PP test for the series after taking the fractional differences of the previous methods

model	BIC	Φ1	Φ2	Θ1	Θ2	MSE	RMSE	MAE
(2, 0.2552436, 0)	2489.26	0.932697	0.008876	0	0	8628.054	92.88732	48.27036
(1, 0.573825, 0)	2498.91	0.158571	0	0	0	9495.135	97.44298	49.21313
(1, 0.0538539, 0)	2475.3	0.947205	0	0	0	8355.203	91.4068	48.19055
(1, 0.4972547, 0)	2497.4	0.16100	0	0	0	9501.976	97.47808	49.58315
(1, 0.49, 0)	2497.23	0.16099	0	0	0	9502.709	97.48184	49.61861
(1, -0.1118070, 0)	2470.63	0.9487734	0	0	0	8053.339	89.7404	46.63232

Comparing the two-stage long memory estimation models with the one-stage long memory estimation model, it turns out that the best model is (1, -0.1118070, 0) with the one-stage estimation method, as it has the lowest BIC criterion as well as the lowest value for MSE, RMSE, MAB.

Table 17. Estimate Model Parameters (1, -0.1118070 , 0)

model	Parameters	Appreciation	p.value
AR(1)	Φ1	0.9487734	0.0000

The mathematical formula for the model (1, -01118070, 0) ARFIMA is as follows:

$$\varepsilon_t = X_t^{-0.1118070} (1 - B) (1 - 0.9487734) \tag{15}$$

6.4. Residual Analysis (1, -0.1118070 , 0) FUZZY_ARFIMA

Residual analysis is considered the most important stage in determining the suitability of the ARFIMA (1, -0.1118070, 0) model for use in forecasting.

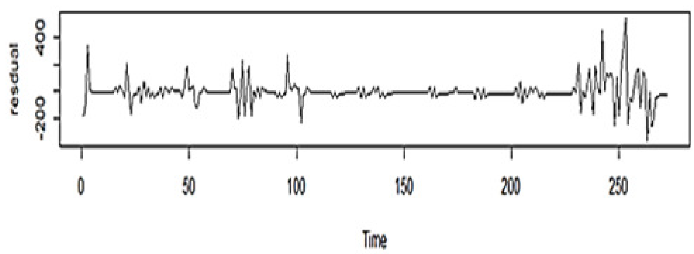

Figure 8. Residual curve for fuzzy series data.

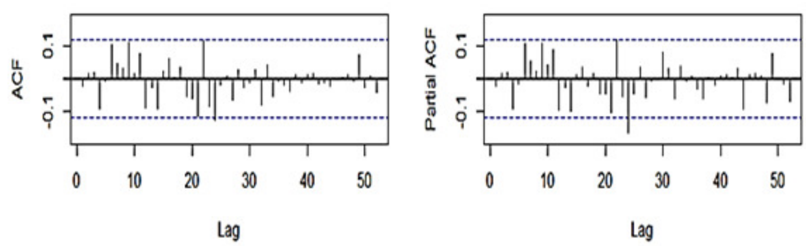

Figure 9. The autocorrelation function of the residuals and the partial autocorrelation function of the residuals.

From the previous figure, we find that all the autocorrelation and partial autocorrelation coefficients for the residuals fall within the confidence limits interval, and this means that the residuals of this model are not related to each other (the resulting errors are random), which indicates that the residuals are stationary.

6.4.1. Ljung box test Below is a Table showing the test results for the residuals of the model (1, 0.2581199, 1):

Table 18. Ljung box test

Model	Q*	p-value
(0, 0.1118070- , 1)	10.524	0.3958

We note that the significant value is p-value > 0.05, which means that the residuals are independent and that the model has passed the diagnosis.

7. Prediction: ARFIMA (1, -0.1118070 , 0) model prediction

After choosing (FUZZY_ARFIMA) as the best model in diagnosis and the best representative model for the data, we used it for prediction. The prediction will be for 12 values and then compare them with the fuzzy values, as shown in the Table below.

Table 19. Predicted values for 12 values

Sequence	Fuzzy-value	Predicted value
1	300	324
2	476	346
3	300	364
4	476	380
5	300	394
6	476	408
7	476	420
8	300	430
9	300	438
10	300	446
11	476	454
12	476	460

8. Steps to build the hybrid model OR ARFIMA- FUZZY

After obtaining the predictive values using the ARFIMA model which represents the linear component $\widehat{L}_t$, we must find the predictive values for the residuals using the FTS model, which represents the nonlinear part $\widehat{N}_t$ to obtain the final predictions of the hybrid model, by performing the following steps:

1. We test the residuals for the ARFIMA model with the BDS test to test the nonlinearity of the residuals, as shown in Table 20:

Table 20. BDS test results for ARFIMA model residuals

Dimensions	BDS	Std. Error	P.value
2	56.0610	3.2494	0.0012
3	112.1221	5.6751	0.0000
4	168.1831	7.3235	0.0000
5	224.2441	7.9897	0.0000

We note that the P.value is less than 0.05 for all dimensional levels, which means that the null hypothesis which states that the model residuals are non-linear is not rejected and therefore these patterns must be analyzed to benefit from them in the prediction process.

2. We take the remainders of the ARFIMA model and run the fuzzy-singh algorithm on them by taking the minimum and maximum limits of the residuals, where $D_{min} = -474$, $D_{max} = 580$, D1= -74, D2=20, since the residuals for the ARFIMA model fall within the interval [-400,600], we divide The remainders are divided into periods of equal length, where the number of periods was (10), as shown in the Table below:

Table 21. Intervals and fuzzy sums for the hybrid model

Periods	Interval values	the group
U_1	[-400,-300]	A_1
U_2	[-300,-200]	A_2
U_3	[-200,-100]	A_3
U_4	[-100,0]	A_4
U_5	[0,100]	A_5
U_6	[100,200]	A_6
U_7	[200,300]	A_7
U_8	[300,400]	A_8
U_9	[400,500]	A_9
U_{10}	[500,600]	A_{10}

3. Below are the sums of fuzzy relationships for the residuals of the ARFIMA model:

Table 22. Intervals and fuzzy sums for the hybrid model

groups	fuzzy second-degree relationship
G1	A1→A8
G2	A2→A2, A3, A4, A5
G3	A3→A1, A4, A5, A6, A8
G4	A4→A2, A3, A4, A5, A6, A9
G5	A5→A2, A3, A4, A5, A6, A7, A8, A10
G6	A6→A2, A3, A4, A5,A6
G7	A7→A3, A4, A5, A7,A8
G8	A8→A4
G9	A9→A7
G10	A10→A6

4. We will fuzzify the values of the residuals for the fuzzy sets to which they belong.
5. After combining the linear component with the non-linear component to obtain the final predictions for the hybrid model according to Equation (4) mentioned previously, the following figure shows the real-time series with predictive values using the hybrid model.

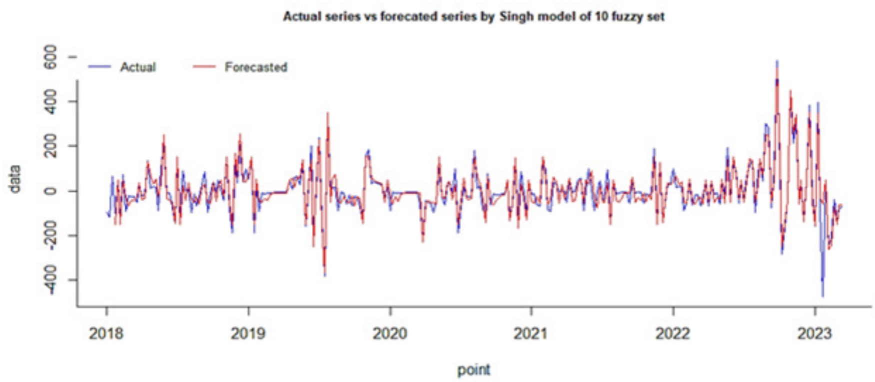

Figure 10. Forecasting actual and predictive time series shapes using ARFIMA-FTS model.

6. By comparing the forecast accuracy standards between the ARFIMA model (1, 0.0179464, 0) and the FTS for the rest of the ARFIMA model, it was found through the research that the hybrid series model ARFIMA_FUZZY(FTS) for the time series under study contains linear and non-linear patterns, and the forecast results for the ARFIMA model were (1, 0.0179464, 0) is more accurate than the hybrid model, and the following Table shows the forecast accuracy standards.

Table 23. Prediction accuracy criteria

Model	MSE	RMSE	MAE
ARFIMA (1, 0.0179464 ,0)	6147.466	78.40578	58.15908
FUZZY_ARFIMA (1, -0.1118070, 0)	8053.339	89.7404	46.63732
ARFIMA-FUZZY (FTS)	1690.287	41.113	28.397

From the Table above, we notice that the ARFIMA-FUZZY(FTS) model is superior to the ARFIMA (1, 0.0179464, 0) model and the FUZZY-ARFIMA model because it is characterized by the lowest value for the forecast accuracy criteria.

9. Conclusions and recommendations

In this research, a new hybrid model was proposed to predict red onion prices in Nineveh Governorate by integrating the linear component of the time series analyzed using the ARFIMA model (1, 0.0179464,0) with nonlinear component analysis using the FTS model for the ARFIMA residuals model (1, 0.0179464). ,0). It has been found through research that ARFIMA hybrid fuzzy time series (FTS) model is effective for predicting the weekly frequency of red onion crop prices. The time series under study are not stationary and contain both linear and nonlinear patterns. Reliability was analyzed by taking the fractional difference d estimated by the EMLE estimator and the Geweke_Porter_Hudak (GPH), Smoothed_priodogram_estimator (dsprio), Fracdiff method, Scaled_Range (R/S), and Whittle_estimator methods. The prediction results using the hybrid model were more accurate than the predictions obtained from the ARFIMA model (1, 0.0179464, 0), and FUZZY_ARFIMA (1, -0.1118070, 0), and the ARFIMA-FTS hybrid model had lower values for these parameters than the ARFIMA model (1, 0.0179464, 0) and FUZZY_ARFIMA (1, -0.1118070,0) so the hybrid model can be used to forecast the future series of red onion prices. It can also be used to forecast time series that contain linear and nonlinear components.

REFERENCES

1. A. M. Abbasov, and M. H. Mamedova, *Application of fuzzy time series to population forecasting*, Vienna University of Technology, vol. 12, pp. 545–552, 2003.
2. H. Abdollahi, and S. B. Ebrahimi, *A new hybrid model for forecasting Brent crude oil price*, Energy, vol. 200, p. 117520, 2020.
3. H. Boubaker, G. Canarella, R.Gupta, and S. M. Miller, *A Hybrid ARFIMA wavelet artificial neural network model for DJIA Index forecasting*, Computational Economicsy, vol. 62, no. 4, pp. 1801-1843, 2023.
4. S. M. Chen, *Forecasting enrollments based on fuzzy time series*, Fuzzy sets and systems, vol. 81, no. 3, pp. 311–319, 1996.
5. S.-M. Chen, and C.-C. Hsu, *A new method to forecast enrollments using fuzzy time series*, International Journal of Applied Science Engineering, vol. 2, no. 3, pp. 234–244, 2004.
6. D. Devianto, K. Ramadani, A. Y. Maiyastri, and M. Yollanda, *The hybrid model of autoregressive integrated moving average and fuzzy time series Markov chain on long-memory data*, Frontiers in Applied Mathematics and Statistics, vol. 8, p. 1045241, 2022.
7. R. Gao, and O. Duru, *Parsimonious fuzzy time series modelling*, Expert Systems with Applications, vol. 156, p. 113447, 2020.
8. C. W. Granger, and R. Joyeux, *An introduction to long-memory time series models and fractional differencing*, Journal of Time Series Analysis, vol. 1, no. 1, pp. 15–29, 1980.
9. J. R. M. Hosking *Fractional Differencing*, Biometrika, vol. 68, no. 1, pp. 165–176, 1981.
10. K. Huarng, *Heuristic models of fuzzy time series for forecasting*, Fuzzy Sets and Systems, vol. 123, no. 3, pp. 369–386, 2001.
11. H. E. Hurst, *Long-term storage capacity of reservoirs*, Transactions of the American Society of Civil Engineers, vol. 116, no. 1, pp. 770–799, 1951.
12. G. Klir, and B. Yuan, *Fuzzy sets and fuzzy logic*, New Jersey: Prentice Hall, 1995.
13. A. W. Lo, *Long-term memory in stock market prices*, Econometrica: Journal of the Econometric Society, vol. 59, no. 5, pp. 1279–1313, 1991.
14. S. Wheelwright, S. Makridakis, and R. J. Hyndman, *Forecasting: methods and applications*, New York: John Wiley & Sons, 1998.
15. J. E. Medina Reyes, S. Cruz Aké, and A. I. Cabrera Llanos, *New hybrid fuzzy time series model: Forecasting the foreign exchange market*, Contaduría y administración, vol. 66, no. 3, pp. 1–23, 2021.
16. S. N. Sivanandam, S. Sumathi, and S. N. Deepa, *Introduction to fuzzy logic using MATLA*, Berlin/Heidelberg, Germany: Springer, 2007.
17. S. R. Singh, *A computational method of forecasting based on fuzzy time seriesI*, Mathematics Computers in Simulation, vol. 79, no. 3, pp. 539–554, 2008.
18. Q. Song, and B. S. Chissom, *Forecasting enrollments with fuzzy time series—Part I*, Fuzzy sets and systems, vol. 54, no. 1, pp. 1–9, 1993.
19. S. A. Tuama, *Using Analysis of Time Series to Forecast numbers of The Patients Malignant Tumors in Anbar Provinc with Malignant Tumors in Anbar Provinc*, AL-Anbar University journal of Economic and Administration Sciences, vol. 4, no. 8, pp. 371–393, 2012.
20. L. A. Zadeh, *Information and control*, Fuzzy Sets, vol. 8, no. 3, pp. 338–353, 1965.