Turkish Journal of Forecasting vol. 06 no. 2 (2022) pp. 83-96

Content list available at JournalPark

Turkish Journal of Forecasting

Journal Homepage: tjforecasting.com

Impact of Structural Break Location on Forecasting Accuracy: Traditional Methods Versus Artificial Neural Network

D. A. Aser[1,*], E. Firuzan[1]

[1]*Dokuz Eylul University, Faculty of Science, Department of Statistics, Tınaztepe Campus, Izmir, Turkey*

A R T I C L E   I N F O

Article history:

Received	15	August	2022
Revision	20	November	2022
Accepted	30	December	2022
Available online	31	December	2022

Keywords:

Structural break
Break location
Conditional heteroscedasticity
Artificial neural network
Forecasting accuracy
Break size

R E S E A R C H   A R T I C L E

A B S T R A C T

Since forecasting future values is fundamental for researchers, investors, practitioners, etc., obtaining accurate predictions is critical in time series analysis. The accuracy is reliant on good modelling and good-quality data. The latter is affected by unusual observations, changes over time, missing data, and structural breaks among others. Economic crises are the major cause of data instability and therefore, this paper focuses on how structural breaks in conditional heteroscedastic financial and macroeconomic data affect forecasting accuracy on short and long-term horizons. More specifically, we are interested in the impact of the location of the structural break and break size on the predictive performance of two linear (ARIMA and Exponential Smoothing) forecasting models and two nonlinear (ARIMA – ARCH and Artificial Neural Network) models. We conducted Monte Carlo simulations and showed that the forecasting accuracy decreases as the structural break location approaches the end of the sample. In addition, break size and length of the horizon show the same impact on the forecasting accuracy as the forecasting error increases with the increase of break magnitude and length of the horizon. We also showed that ARIMA – ARCH model is the best performing in the absence of a structural break while the artificial neural network model outperforms all the competing models in the presence of structural break, especially in large break sizes and long horizons. Last, we applied the above techniques to forecasting daily close prices of brent oil and Turkish Lira – USD exchange rates out–of–sample, and similar results were found.

1. Introduction

Time series models are used for various purposes, including forecasting future outcomes, analysing past behaviours, and making policy recommendations. In this regard, we usually experience unexpected changes (breaks) in model parameters in financial time series data, perhaps due to policy changes, technological advancement, financial crises, and natural disasters. Events like the 2007–2008 economic crisis and the COVID-19 issue are good examples of structural break causes.

Numerous studies revealed that several crucial and mainly used economic indicators have structural breaks. For instance, Stock and Watson [1] used several standard statistical tests to examine 76 monthly U.S. economic time series relationships for model instability. The series examined included key economic indicators such as interest rates, stock prices, industrial production, and consumer expectations. Their findings showed that a significant

*Corresponding author.
E-mail addresses: caseyr025@gmail.com (Daud Ali Aser)

https://doi.org/10.34110/forecasting.1162548

proportion of the indicators have structural breaks. Moreover, Pettenuzzo and Timmermann [2] investigated the predictability of stock returns and asset allocation in the presence of structural breaks, arguing that structural breaks are most likely when the parameters of return prediction models are estimated on data samples spanning several decades. Many other authors testified the presence of structural breaks in economic and financial time series, among them Koop and Potter[3], Siliverstovs and Van Dijk[4].

Structural breaks reduce forecasting accuracy and consequently misleads policy recommendations and other prediction purposes. Clements and Hendry [5] stated that macroeconomic models cannot forecast well because of the structural breaks and warned that ignoring the breaks leads to a wide discrepancy between theory and practice. Hansen [6] also argued that it is perilous to ignore structural breaks in economic time series because forecasts may be inaccurate and therefore misleads policy implications. Yin [7] considered how to forecast a time series variable of interest in the case of parameter instability by emphasizing model selection criteria. In this paper, several break sizes, different test sample sizes, and various data-generating designs are discussed using simulation evidence and empirical analysis. All the experiments show that cross-validation weights outperform the competing methods in the under of structural breaks and heteroscedasticity.

Altansukh and Osborn [8] conducted a study on forecasting time series using structural break inferences. This study suggested averaging forecasts over sub-samples indicated by a confidence interval or set for the break date instead of relying on a point estimate of a coefficient break date. Furthermore, they investigated whether explicit consideration of a potential variance break and using a two-step methodology improves forecast accuracy compared to robust heteroskedasticity inference. Their Monte Carlo results and empirical implementation of U.S. productivity growth show that averaging with the likelihood ratio-based confidence set typically outperforms other methods. Moreover, Bouri et al. [9] examined how U.S. ethanol prices and volatility dynamics can be forecasted under structural breaks. The findings of this study revealed that ARCH/ GARCH models incorporating the breaks improve the prediction of U.S. ethanol market volatility. Furthermore, volatility persistence decreases when structural breaks are included in the GARCH models. Their findings suggested that ignoring such breaks could mislead the U.S. biofuel industry's risk assessment procedure.

Bauwens et al. [10] also joined the other authors by revealing that ignoring change points in financial time series produces poor forecasts and often gives the spurious impression of nearly integrated behaviour of the time series. In this regard, they contributed with an algorithm based on the Markov chain Monte Carlo (MCMC) method for Bayesian inference in AR-GARCH models subject to an unknown number of structural breaks at unknown dates. This method treats break dates as parameters and determines the number of breaks using the marginal likelihood criterion. This study also considered both pure-change point and recurrent regime specifications and showed how to forecast in the presence of structural breaks. The algorithm's efficiency is illustrated through simulations and empirical analysis on eight financial time series of daily returns from 1987-2011. Kaushik et al. [11] evaluated the performance of several deep neural networks (DNN) in forecasting time series with multiple structural breaks and high volatility. The models they considered here were Multi-layer Perceptron (MLP), Convolutional Neural Network (CNN), and RNN with Long Short-Term Memory (LSTM-RNN) and RNN with Gated-Recurrent Unit (GRU-RNN). These models were implemented on 10 Indian financial stocks data to forecast single-step and multi-steps. The DNN methods demonstrated convincing performance for single-step forecasting, while the performance of these methods showed that long forecast periods have a negative effect on performance.

Despite the extensive literature on the structural break and its impact on forecasting, little emphasis is paid to how the location of the structural break in time series affects predicting accuracy. Questions such as whether forecasting accuracy decreases as the structural break approaches the end of the series remain unaddressed.

Another issue we usually experience in time series is the problem of conditional heteroscedasticity. Assuming conditional homoscedasticity, especially in the case of structural breaks in forecasting macroeconomic conditions or financial time series may seem restrictive. For this reason, this paper investigates and compares the predictive ability of four (linear and nonlinear) models including Exponential Smoothing (ETS), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average -Autoregressive Conditional Heteroskedastic (ARIMA – ARCH) and Artificial Neural Network (ANN) models in out-of-sample forecast setting in the presence of structural breaks and ARCH innovations based on simulations evidence and empirical analysis. The rest of the paper is structured as follows. Section 2 explains the methods used for structural break tests, forecasting, and forecasting accuracy measures. Section 3 describes the data generating process for simulation. Section 4 presents simulation results, section 5 contains empirical results while section 6 provides conclusions.

2. Methodology

The forecasting methods used in this paper include Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average -Autoregressive Conditional Heteroskedastic (ARIMA – ARCH) and Artificial Neural Network (ANN). Chow Test, CUSUM Test and Bai-Perron test were used to detect the presence of structural break/s, while Engle's Lagrange Multiplier tests was introduced to detect the presence of ARCH effect. Lastly, Root Mean Squared Forecast Error (RMSE) is used to measure the forecasts accuracy.

2.1. Structural break testing methods

Identifying structural breaks in time-series modelling is essential in statistical and econometric literature. By considering the statistical characteristics of structural breaks and time series data, this study implemented the Chow, CUSUM, and Bai and Perron tests. The Chow [12] test belongs to a class of structural change tests based on F statistics. It is commonly used in time series analysis to test for the presence of a single structural break at a known period (for example, a significant historical event such as a war or pandemic). The main disadvantage of the Chow test is that the change point must be specified in advance; however, Supremum Test and Wald and Lagrange Multiplier test statistics solved this problem. These are tests based on Chow statistics and do not require explicitly stating a particular change point. The CUSUM test developed in Brown et al.'s seminal paper [13], a widely used test for model parameters stability, contains cumulative sums of standardized residuals. This test comes under the class of generalized fluctuation test framework of structural change tests. Bain and Perron’s [14] test is also utilized to test if multiple structural breaks are present and estimate break dates.

2.2. Forecasting models

This study employs two linear (ARIMA and Exponential Smoothing) forecasting models and two nonlinear (ARIMA – ARCH and Artificial Neural Network) models in an out-of-sample forecast setting.

2.2.1. Exponential smoothing (ETS) model

The exponential smoothing class of models was proposed in 1950s [15], [16] and they are capable of producing time series forecasts by employing a weighted average of its historical values and allocating more weight to recent observations. These models consist of observation/ measurement equation which denotes the observed data and state equations which stands for states (level, trend, seasonal) change over time and that is why are called state-space models. Exponential smoothing models are labelled with the notation ETS for the terms of Error, Trend, Seasonal. The ETS models have got variations because of the trend and seasonal combinations. The trend component may be none (N), may be additive (A), or additive damped (Ad) while the seasonal component can be none (N), additive (A) and multiplicative (M) yielding nine variants of exponential smoothing models [17].

The component form of $ETS(A, N, N)$ model with additive errors is as following:

Forecast equation $$\hat{y}_{t+1|t} = \ell_t \tag{1}$$

Smoothing equation $$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \tag{2}$$

By re-arranging the smoothing equation, we get the

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \ell_{t-1} + \alpha e_t \tag{3}$$

where $\ell_t$is the estimated level, $\hat{y}_{t+1|t}$is one step-ahead prediction for time $t + 1$ which results from the weighted average of all historical data while $0 \leq \alpha \leq 1$ is the smoothing parameter and $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ is error at time t.

Other ETS models can be written in similar fashion for each of the exponential smoothing methods. An automated selection procedure is utilized to identify the exponential smoothing models by using “ets” function from “$forecast$” package in R environment proposed by Hyndman et al. [18]. This function automatically identifies which model best suits the given time series, estimates the model parameters, and returns information about the fitted model. It can use all information criteria but the bias-corrected Akaike criterion (AICc) is the default information criterion to select an appropriate model.

2.2.2.Autoregressive integrated moving average (ARIMA) model

One of the most extensively used models for time series forecasts is the Autoregressive Integrated Moving Average (ARIMA) model, which was first suggested by Box and Jenkins [19] . The ARIMA model is a generalization of an ARMA model. ARIMA(p,d,q) is a non-seasonal ARIMA model with non-negative parameters p, d, and q, where p is the number of time lags referred to as the order of the autoregressive model, d is the degree of differencing, and q is the order of the moving-average model.

The ARIMA models transform a non-stationary series to a stationary series through a sequence of differencing steps. A time series y_t is integrated of order d, if $\nabla^d y_t$ is stationary and

$$\nabla y_t = y_t - y_{t-1} \tag{4}$$

where y is the time series and t is the time index.

After we transform the time series into stationary, the estimation is done like the following:

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t \tag{5}$$

where ϕ_i denotes the coefficients of AR terms of order p and θ_i denotes the coefficients of the MA terms of order q and ε is a random term that is white noise and μ is a constant term. The model parameters are estimated using Maximum Likelihood Estimation (MLE).

For best ARIMA model selection, we relied on $auto.arima$ function from "$forecast$" package in R proposed by Hyndman and Khandakar [20]. This function identifies best ARIMA model based on either the Akaike information criterion (AIC), the bias-corrected Akaike criterion (AICc) and the Bayesian information criterion (BIC) value. This function searches over possible models within the order constraints provided.

2.2.3.Combined ARIMA and ARCH (ARIMA – ARCH) model

The error structures in ARIMA models are assumed to be generated from linear combinations of the innovations, with the innovations being, *i.i.d.* normal random variables, but that is not always the case. The variance fluctuates widely with time in many financial and economic time series, and ARIMA models cannot capture this behaviour. To address these conditions, ARCH (autoregressive conditionally heteroscedastic) proposed by Engle [21] and GARCH (generalized ARCH) by Bollerslev [22] models have been constructed. For this reason, the ARIMA and GARCH models are combined. While the ARIMA part of the model estimates the conditional mean, a GARCH part of the model estimates the conditional variance included in the residuals of the ARIMA estimation. The estimation based on the combined model is as follows:

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \varepsilon_t \tag{6}$$

$$\varepsilon_t = \sigma_t a_t \tag{7}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \tag{8}$$

where a_t is white noise ω and α_1 are the parameters of the model. Substituting for σ_t^2, we get:

$$\varepsilon_t = a_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2} \tag{9}$$

Then ε_t is ARCH (1) model which stands for autoregressive conditionally heteroscedastic. An ARCH(q) process is given by:

$$\varepsilon_t = a_t \sqrt{\omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2} \tag{10}$$

GARCH(q,p) can be formulated if we define $\varepsilon_t = \sigma_t a_t$ where a_t white noise and σ_t^2 is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{11}$$

where α_i and β_j are parameters of the model. The forecast is calculated by combining the output of the ARIMA part and the ARCH/GARCH part of the model.

2.2.4. Artificial neural network (ANN) model

The neural network connects the predictors and the outcome through a series of layers. In each layer, some operation is performed on the input data to produce an output. The output of one layer becomes the input for the next layer. As shown in Figure 1, the structure of ANN is characterized by three layers:

1. An *input layer* that accepts the input values of the predictors or lagged terms in the time series case.
2. *Hidden layers* that generate derived variables. A hidden layer receives inputs from the preceding layer, which is the input layer, performs some calculations on those inputs based on some activation function, and creates an output.
3. An *output layer* receives input from the latest hidden layer and predicts future values.

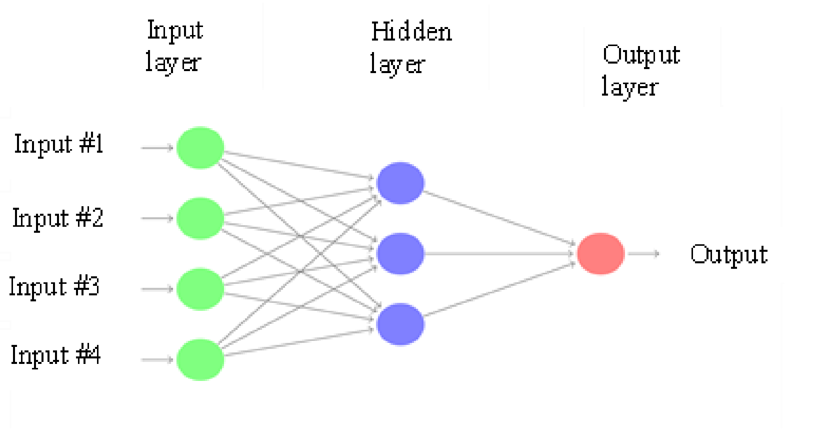

Figure 1. Artificial neural networks architecture

Each layer contains nodes denoted by circles. While input layer nodes are the original predictors, hidden layer nodes are the weighted sum of the inputs to which some monotone function, called an activation function, is applied. Common functions are linear, exponential, and s-shaped functions such as the logit and hyperbolic tangent or, in other words, sigmoid functions.

In time series approach, the relationship between the output ((y_t) and the inputs ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) can be represented by Equation 12:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_{jg} \left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i} \right) + \varepsilon_t \tag{12}$$

where α_j ($j = 0,1,2, \dots, q$) and β_{ij} ($i = 0,1,2, \dots, p; j = 1,2, \dots, q$) are the model parameters often called the connection weights; p is the number of input nodes and q is the number of hidden nodes. Logit activation function is used to map the input values into the hidden layer and from the nodes of the hidden layer to the output, linear function is used as the series takes values on a continuous range. The logistic function is as given in equation 13.

$$g(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

Hence, the ANN model performs mapping from past observations $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ to the future y_t, i.e.,

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) + \varepsilon_t \tag{14}$$

If the lagged values of the time series are used as inputs to a neural network just as used lagged values in a linear autoregression model, it is called a neural network autoregression or NNAR model. In this paper, we only consider feed-forward networks based on $NNAR(p, k)$ where p indicates lagged inputs and k nodes in the hidden layer [17].

Neural network autoregression models were identified using "$nnetar$" function from "$forecast$" package in R produced by Hyndman [23]). This function automatically fits a neural network model to the given time series with lagged values of the series as inputs, so it is a nonlinear autoregressive model and that is why we considered this method as it ables to capture the non – linearity int the financial data.

2.3. Forecasting accuracy metrics

To compare the forecast methods and assess the accuracy of our out-of-sample forecasts against the reserved observed values in the evaluation sample termed as the test set, we use Root Mean Squared Forecast Error (RMSE) computed as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (\hat{y}_{t+h|t} - y_{t+h})^2}{T}} \tag{12}$$

where $\hat{y}_{t+h|t} - y_{t+h}$ states forecast error and training data given by $\{y_1, \dots, y_t\}$ and the test data is given by $\{y_{t+1}, y_{t+2}, \dots\}$.

Root mean squared error (RMSE) is widely used in comparing forecast methods applied to an individual time series, or to numerous time series with the same units[17]. Since our time series data simulations are replicated based on the same data generation process, RMSE is an appropriate measure to compare predictive performance of the considered forecast methods.

3. Data Generating Process

In generating data, it is considered the following design of AR(2) process with autoregressive conditionally heteroscedastic errors ARCH(2),

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{13}$$

$$\varepsilon_t = \sigma_t a_t \tag{14}$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 \tag{15}$$

where a_t is white noise for ω, α_1 and α_2 are the parameters of the variance model. Substituting for σ_t^2 we receive:

$$\varepsilon_t = a_t \sqrt{\omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2} \tag{16}$$

with the help of equation (19), $AR(2) + ARCH(2)$ process can be generated.

The parameters of both AR part and ARCH part of the data generating process mentioned above were specified as: $\mu = 5, \phi_1 = 0.65, \phi_2 = 0.25, \omega = 0.2, \alpha_1 = 0.45, \alpha_2 = 0.3$. These values were chosen arbitrarily. The graphical visualization of the simulated data is shown in figure 2 for illustration purposes.

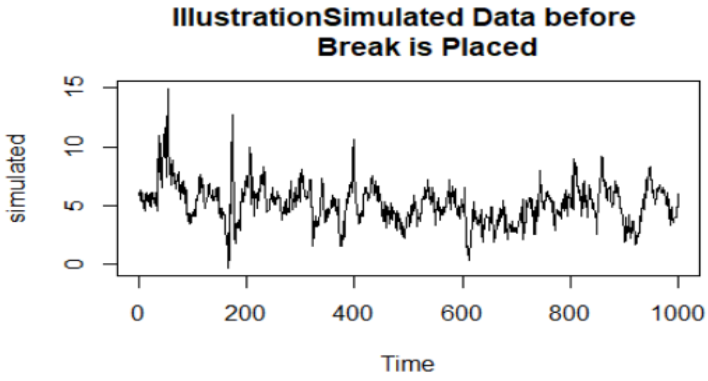

Figure 2. Illustration of simulated data structural break is placed

To examine the effect of structural break location on forecasting accuracy, we first consider the case (case I) of forecasting with the stable data, i.e., before the structural break is placed. We then construct three unstable cases by allowing a one-time permanent structural break in the mean of the series in each case. The breaks were positioned at 25%th observation termed as break at first quartile(Q1), 50%th observation as second quartile (Q2) and 75%th observation as third quartile (Q3) of the training sample (n). We let the structural break take the multiplicative form, which means that if the pre-break mean is μ, then the post-break value is $\delta\mu$, where δ is a break size parameter. In each situation we considered three different break sizes namely break of small magnitude (δ = 3), medium break size (δ = 5) and large break size (δ = 10). For illustration purpose of the location of the structural break and break sizes see figure 3.

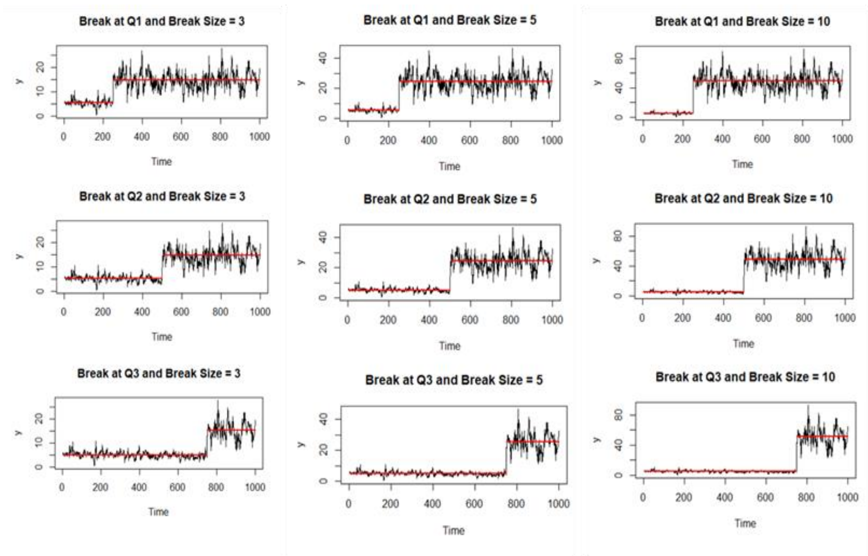

Figure 3. Illustration of break locations and break sizes of simulated data

The total sample size, $T = n + h$ is 1000. To investigate if the choice of the horizon size has an impact on forecasting results, we reserve the first 990, 995 and 997 (n = 990, n = 994 and n = 997) observations as the training sample and the rest as the prediction sample (h = 10, h = 6 and h = 3) in three separate experiments in each case. The simulation process is conducted with 5000 times replications. The hyperparameter values of the ANN based on RMSE metric are given in Table 1.

Table 1. Hyperparameter values

Break Size	No. of optimal lags (inputs)	No. of hidden layers	No. neurons in the hidden layer	Model structure notation
No break	4	1	3	NNAR (4,3)
(δ) = 3	8	1	5	NNAR (8,5)
(δ) = 5	9	1	5	NNAR (9,5)
(δ) = 10	6	1	4	NNAR (6,4)

In each case, to evaluate and compare predictive performance, we generate forecasts using four methods: Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing, ARIMA ─ ARCH, and Artificial Neural Network. Root mean squared forecast error (RMSE) is used to assess their predictive performance.

4. Simulation Results

Simulations results are presented in four cases. The first case is the stable data where there is no structural break while the other three cases are based on the break size and in each of them three break locations are considered.

Case I: Forecasts stable simulated data

Simulation results for the stable case are shown in Table 2. We can see from Table 2 that all the models are performing well when there is no structural break as the RMSE values are small. It can also be seen that the combined ARIMA and ARCH (ARCH) model outperforms all the classical ARIMA, Exponential Smoothing (Exp), and even the Artificial Neural Network (ANN) models across all considered test horizons. At the same time, ANN is the next best model.

Table 2. RMSE values of stable simulated data

Model	h = 3	h = 6	h = 10
ARCH	0.773	0.941	1.083
ARIMA	0.791	0.975	1.147
Exp	0.818	1.017	1.216
ANN	0.785	0.961	1.108

The ARIMA and exponential smoothing methods are less performing models in this case. This was expected because they failed to capture the nonlinearity of the process under consideration. RMSEs increase monotonically as the horizon size increases in all models.

Case II: forecasts when break size (δ) = 3

Simulation results for case II where the structural break size is relatively small (δ = 3) are reported in Table 3. It can be seen from the table that the RMSEs of all the models for all test horizons increased in comparison with RMSEs in Table 2 of stable data. It is also clear that ANN outperforms all the other models across all considered test horizons if the break is in the first quartile (Q1) or in the second quartile (Q2). ARIMA takes over the ANN when the structural break is in the third quartile (Q3). Moreover, RMSEs increase monotonically as the horizon size increases in all models. RMSEs also slightly increase as the break moves from Q1 through Q2 to Q3.

Table 3. RMSE values when break size (δ) = 3

Break size (δ) = 3	Model/Horizon	h = 3	h = 6	h = 10
	Structural Break at Q1			
	ARCH	2.411	2.990	3.559
	ARIMA	2.407	2.963	3.501
	Exp	2.460	3.058	3.654
	ANN	2.380	2.905	3.392
	Structural Break at Q2			
	ARCH	2.419	3.010	3.597
	ARIMA	2.414	2.982	3.520
	Exp	2.465	3.060	3.661
	ANN	2.405	2.946	3.454
	Structural Break at Q3			
	ARCH	2.461	3.106	3.786
	ARIMA	2.432	3.033	3.591
	Exp	2.459	3.104	3.693
	ANN	2.464	3.071	3.646

Case III: Forecasts when break size (δ) = 5

Simulation results for case III where the structural break size is relatively medium (δ = 5) are presented in Table 4.

Table 4. RMSE values when break Size (δ) = 5

Break size (δ) = 5	Model/Horizon	h = 3	h = 6	h = 10
	Structural Break at Q1			
	ARCH	4.041	5.024	5.996
	ARIMA	4.008	4.941	5.841
	Exp	4.099	5.094	6.089
	ANN	3.966	4.830	5.661
	Structural Break at Q2			
	ARCH	4.056	5.086	6.074
	ARIMA	4.036	5.032	5.944
	Exp	4.087	5.164	6.142
	ANN	4.042	5.016	5.859
	Structural Break at Q3			
	ARCH	4.150	5.237	6.401
	ARIMA	4.080	5.095	6.035
	Exp	4.104	5.173	6.158
	ANN	4.090	5.086	5.982

RMSEs increase monotonically with the increase of break magnitude and horizon size in all models. RMSEs also marginally increase as the break moves from the first quartile through the second quartile to the third quartile. For horizons 6 and 10, ANN is shown to be the best choice wherever the break is located. But for horizon 3, ANN is best if the break is at Q1, while if the break is at Q2 and Q3, ARIMA gives better results than ANN.

Case IV: Forecasts when break size (δ) = 10

Simulation results for case IV where the structural break size is relatively large (δ = 10) are reported in Table 5. Same to Case II and III, RMSEs increase with the break size, horizon size, and the location of the structural break.

Table 5. RMSE values when break size (δ) = 10

Break size (δ) = 10	Model/Horizon	h = 3	h = 6	h = 10
	Structural Break at Q1			
	ARCH	8.101	10.086	12.047
	ARIMA	8.020	9.890	11.719
	Exp	8.177	10.181	12.171
	ANN	7.936	9.626	11.284
	Structural Break at Q2			
	ARCH	8.179	10.272	12.290
	ARIMA	8.103	10.141	11.951
	Exp	8.211	10.358	12.318
	ANN	8.067	9.927	11.583
	Structural Break at Q3			
	ARCH	8.504	11.137	13.465
	ARIMA	8.193	10.445	12.148
	Exp	8.228	10.640	12.365
	ANN	8.170	10.378	12.012

On the other hand, regardless of the horizon size and the location of the break, ANN is the best performing model among the applied models when the break size is large (δ = 10).

Summary of the Simulation Results

In the simulation part of this paper, we examined how the structural break affects forecasting accuracy. In doing so we investigated three different test locations (Q1, Q2, Q3) for the structural break, three break sizes ($\delta = 3$, $\delta = 5$, $\delta = 10$), and three forecast horizons (h = 3, h = 6, h = 10). Root mean squared forecasting error (RMSE) is used for the comparison tool. It is found that:

1. The results of the stable case give good guidance for researchers showing how is difficult to analyse financial and economic time series despite the non-existence of structural break in the series. Table 2 shows the performance of forecasting with stable data. According to the results in Table 2, the ARIMA and Exponential models are not performing well. It can be explained by the fact that the data under study is generated from a nonlinear process. Since the asymmetry and nonlinearity properties exist naturally in financial and economic time series data regardless of the presence of structural break, it is proposed that ARIMA and Exponential Smoothing models should be used cautiously in nonlinear processes.
2. For the case of a relatively small break magnitude ($\delta = 3$), ANN is superior to all the other models across all considered test horizons if the break is in Q1 or Q2. ANN loses its dominance and ARIMA overtakes when the structural break is in Q3. The failure of ANN when the break size is too small and the structural break is in the third quartile can be explained as ANN models associated with the testing part of the series [24], the effect of structural break which is in the testing part may have further negative impact on the performance of ANN. This could be the first reason for this result. The shifting of the series because of the structural break in the last quartile, the nonlinear behaviour of economic time series could turn over to linear behaviour. Hence, ARIMA generates better performance than ANN, especially when the researchers deal with small size structural break. This could be the second reason for result 2.
3. In the case of comparatively medium break size ($\delta = 5$) in Table 4, ANN is the best for horizons (h = 6) and (h = 10) wherever the break is located. For horizon (h = 3), ANN is still the best if the break is at Q1, while ARIMA gives better results than ANN if the break is at Q2 and Q3. Although ANN models are associated with the testing part of the series, it is not affected by the location of the structural break especially for Q3 which its size is 5. On the contrary the case of breaks size 3, ANN recovers itself in the long-term horizon. On the other hand, ARIMA is a univariate model and is also known as a method that gives very successful results in short-term forecasts [24]. ARIMA models need the stability of the data as it requires basic statistical properties such as mean, variance, covariance, or autocorrelation to be constant over time periods [25]. So, the behavior of the financial time series may change around the statistical properties (mean, variance etc.) because of the larger break size of series. Therefore, the performance of ARIMA for the long-term horizons is not good as the short-term horizon in the presence of structural breaks. This could mean that ANN gives better results for the longer horizons, but ARIMA’s performance is getting worse as break size and horizon increase. From the point of economical view, making forecast for the long term is more valuable but riskier. This is because the economic time series data has dynamic properties and may have been affected by sudden changes, so the researcher must keep in mind those bad events. As a result, researchers should prefer ANN for the larger size structural break in the series.
4. In the case of a relatively large break size ($\delta = 10$), regardless of the horizon size and the location of the break, ANN is the best-performing model among the applied models. The results for largest break size ($\delta = 10$) supported the interpretations of relatively smaller break size ($\delta = 5$).
5. RMSEs increase with horizon length, break size, and the break location moving from the first quartile, second quartile to the third quartile of the data. All the models have the smallest RMSEs when the horizon is short (h = 3), the break size is small ($\delta = 3$) and the break is located at the beginning of the training set (Q1). The largest RMSEs are found when the horizon is long (h = 10), the break size is large ($\delta = 10$) and the break locates close to the end of the training set (Q3).

Our proposal here is:

When the break size is 3 (δ=3, around the mean) and in the last quartile, ARIMA, which is a mean-based method, should be preferred in comparison with ANN. When the break size is medium as (δ=5), and the break locates in Q2 and Q3, ARIMA should be used for the short-term horizons. Except conditions which ARIMA is performing well, we propose using the ANN in any other case.

5. Empirical Applications

In addition to the simulation evidence, we applied our considered methods to forecast the daily Turkish Lira against US Dollar (USD - TL) exchange rate and Brent oil prices. We selected these two financial data because they exhibit typical structural breaks we are going to investigate in this study. We extracted data by windowing according to the time periods where structural breaks were found. It is extracted in such a way that the break location fits the interested test locations (Q1, Q2, Q3) in line with the simulations section.

Both Brent Oil Prices and the Turkish Lira - US Dollar exchange rate experienced sudden shocks and fluctuations in recent years, making it challenging to forecast their future values. The Close prices of Brent Oil ranging from (20.09.2017 – 13.12.2019) are accessed through the *quantmod* package in R. There is a sharp jump peaking on October 3, 2018, which can be considered a change point that suddenly reverses from the rapid rise, or as a reflection of economic practice. The close prices of the USD - TL exchange rate in the range (9.02.2017 – 26.02.2020) are extracted from Yahoo Finance. Likewise, in the USD - TL exchange rate data, there is a change point on August 10, 2018, due to the crisis and depreciation of the Turkish Lira against the US Dollar. To be consistent with what we have done in the simulations section, we drew three windows of size 400 each for Brent Oil and three windows of size 500 each for TL – USD to represent the three structural break test locations (Q1, Q2, Q3) as illustrated in figure 4.

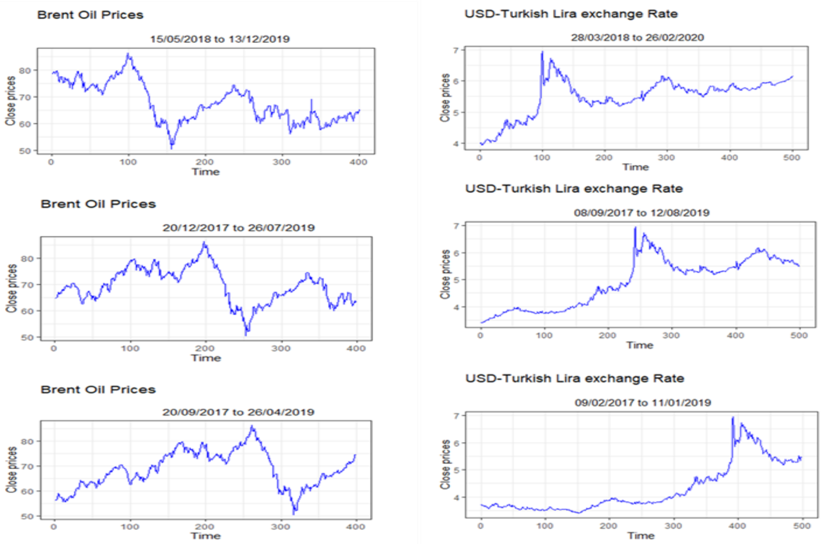

Figure 4. Brent oil prices and TL – USD exchange rate

The left part of figure 4 illustrates Brent Oil Prices while the right part shows the TL – USD exchange rates. While both data sets exhibit at least one structural break in each case, crude oil prices show much volatility greater than the exchange rate.

Structural Break Testing

Prior to fitting and forecasting with the models under consideration, we first check the presence of structural breaks in our data and estimate the break date. Bai and Perron test [14] made possible to estimate the break date and the intercept of each segment and that is done with the help of the R function breakpoints from the strucchange package. Table 6 shows the estimated break date, segments, and intercepts in both brent oil and USD – TL exchange rate data. The presence of structural break is also tested using Chow test and CUSUM test and the results are presented in Table 7.

Table 6. Break dates and segmentation using Bai and Perron test

Data	Break Date	Segment	Intercept
USD - TL Exchange Rate	10/7/2018	09/02/2017 - 10/07/2018	3.817
		11/07/2018 - 26/02/2020	5.696
Brent Oil Prices	9/11/2018	20/09/2017 - 09/11/2018	70.519
		12/11/2018 - 13/12/2019	63.525

Table 7. Significance of the structural break

Data	Test Statistic		P-value
USD - TL Exchange Rate	Chow (supF) test	39.661	8.774e-08
	CUSUM Test	13.185	2.2e-16
Brent Oil Prices	Chow (supF) test	22.098	0.0004
	CUSUM Test	5.4359	2.2e-16

The test statistics and p-values results of Chow test and CUSUM test in Table 7 show that there is significant structural break. As can be seen in Table 6, USD - TL Exchange Rate contain structural break in 10/7/2018 and estimated intercept in the segment of 9/2/2017 - 10/7/2018 is 3.817 while the segment 11/7/2018 - 26/2/2020 should have 5.696 as intercept. Brent Oil Prices also contain structural break at 9/11/2018. The estimated intercepts for 20/9/2017 - 9/11/2018 and 12/11/2018 - 13/12/2019 segments are 70.519 and 63.525 respectively.

Forecasts Accuracy Evaluation in Empirical Analysis

Like simulations, we employed the four forecasting methods we used in simulated data and then forecasted out–of–sample and calculated root mean squared error (RMSE). We also chose evaluation sample sizes same as those in simulations (h = 3, h = 6, h = 10). The smaller the value of RMSE for a given method compared to the other methods the better that method is performing. The empirical results of both considered data are presented in Table 8 side by side.

Table 8. RMSE values of empirical analysis

Brent Crude Oil Price				USD - TL Exchange Rate			
Structural Break at Q1				Structural Break at Q1			
Model/Horizon	**h = 3**	**h = 6**	**h = 10**	**Model/Horizon**	**h = 3**	**h = 6**	**h = 10**
ARCH	0.2573	0.8358	1.1164	ARCH	0.0256	0.0635	0.0725
ARIMA	0.3074	1.1437	1.5038	ARIMA	0.0314	0.0653	0.079
EXP	0.3072	1.1440	1.5034	EXP	0.0272	0.0625	0.0733
ANN	0.2342	0.816	1.017	ANN	0.0259	0.0511	0.0571
Structural Break at Q2				Structural Break at Q2			
ARCH	0.4911	1.3326	1.9648	ARCH	0.0370	0.0919	0.1165
ARIMA	0.5308	1.4246	1.9287	ARIMA	0.0335	0.0892	0.1255
EXP	0.4434	1.2377	1.9873	EXP	0.0388	0.0910	0.1176
ANN	0.347	1.1628	1.9339	ANN	0.0282	0.0708	0.0739
Structural Break at Q3				Structural Break at Q3			
ARCH	1.3475	2.2675	2.4315	ARCH	0.0341	0.1509	0.1693
ARIMA	1.6042	2.1624	2.2808	ARIMA	0.0379	0.1008	0.132
EXP	1.3213	2.2603	2.2804	EXP	0.0813	0.1045	0.1333
ANN	1.3109	2.1177	2.1115	ANN	0.0295	0.0994	0.0887

i. Brent oil price continuously fluctuates, causing instability in the economy. Oil prices' sudden ups and downs have direct or indirect impacts on economics since they produce specific issues in global economics. Due to this rapid variability, predicting oil prices has global importance; however, the bevahiour of brent oil prices is unstable and difficult to predict. Among the considered model for Brent Oil Prices, ANN outperforms the others, including the ARCH model, in all horizons, no matter where the break locates (Q1, Q2, Q3). Even though Oil prices are much more volatile, ANN is more resilient to variance heterogeneity and structural break than the other forecasting techniques in consideration. Imbalanced data, which has nonlinear trend, changeable cyclical fluctuations, and high volatility, is a big issue in time series analysis. It should also be noted that the existence of high volatility may mask the effect of structural break on the performance as well. ANN should be the first preferable methodology for complex economical or financial data.

ii. After 2017, there is a rapid fluctuation in Turkish Lira value against Dollar. The Turkish US Dollar – Turkish Lira exchange rate series has more statistically stable properties in comparison with brent oil price series. The series has trend component and less volatility. In this sense, the location of the structural break is important for the data. If the structural break is in Q1, the series seems more volatile than the structural break in Q2 and Q3. For USD - TL exchange rate, ANN still gives the smallest RMSEs in all the cases except the short horizon (h = 3) when the break is in Q1. In this case, the ARCH model takes the lead. Hence, the ARCH model may be preferred for short–term forecasts of less volatile data when the break is at the very beginning of the training set (Q1). The reason could be that the variance spectrum of the data is broadened because of the location and size of the structural break, and the volatility becomes more effective. Therefore, the ARCH model in Q1 for the short-term horizon generates less RMSE compared to ANN.

iii. All models taken into account here are affected by the location of the break as RMSEs are smallest when the break locates at Q1 and largest when the break is in Q3. This result is true for exchange rate and oil price data and all applied horizons.

iv. For Crude oil prices data, which is relatively more volatile, RMSEs increase faster as the break location progresses from Q2 to Q3 than its progress from Q1 to Q2 for short horizons (h = 3, h = 6), and it is the opposite for long horizons (h = 10). This is true for all models used here.

v. For the exchange rate data, which is relatively less volatile, the only agreement between the models is that RMSEs increase as the break location moves from Q1 to Q2, which is not the case in Q3. There is no considerable agreement between the used models regarding the speed of RMSEs increase.

Our suggestion is:

For the imbalanced and complex economic and financial data, ANN should be preferred in any case regardless of location and size of the break. The dynamical behaviour of the series before or after structural break does not affect the success of ANN. In fact, just in the Q1 case, which is covered by the training set, structural break has caused high volatility, affecting ANN negatively in this case. If structural break causes volatility in economic data or the structural break and volatility overlapped in time and the break is in the first quartile, ARCH model should be used.

6. Conclusion

The aim of this paper was to answer the question of how a structural break affects forecasting performance. More specifically, we are interested in the impact of break location and break size on the predictive performance of competing methods (ARIMA, Exponential, ARIMA - ARCH and Artificial Neural Network) in various simulation cases. Last, we applied the above techniques to forecasting daily close prices of brent oil and TL – USD exchange rates out – of – sample and compare the performance of the applied models.

The study found that apart from the impact of structural break on the forecasting accuracy, break location and break size have implications on prediction. This result is found both in simulations and empirical analysis. Forecasting error increases with the increase of break magnitude and the location of the structural break. The error is minimum when the break is in the first quartile, it is more when the break is in the second quartile, and it is the most when break locates in the third quartile.

Comparing the competing models, ARCH model is the best performing model in the absence of structural break while ANN is the next in performance. For the cases of the various structural break sizes and break locations, ANN outperforms all the other models with the exception of few places. ANN also shown superiority in the empirical analysis in most of the cases under investigation. Hence, while structural break affects the performance of all the examined models, ANN is the most resilient for the breaks and instability of the parameters.

The authors would suggest based on the findings of this study that other sophisticated forecasting methods such as deep learning and forecast combination methods should be used to study deeply the impact of the structural break on forecasting accuracy. More specifically, the effect gets worse when the break is in the third quartile of the sample, and it is recommended that this area to be scrutinized intensely.

References

[1] J.H. Stock, M.W. Watson, Evidence on Structural Instability in Macroeconomic Time Series Relations, Journal of Business & Economic Statistics, 14 (1996), 11–30.

[2] D. Pettenuzzo, A.Timmermann, Predictability of Stock Returns and Asset Allocation under Structural Breaks ∗, Journal of Econometrics, (2010). pp. 60–78. doi: 10.1016/j.jeconom.2011.02.019.

[3] G. Koop, S. Potter, Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series?, SSRN Electron. J. 4 (2005) 37–55. doi:10.2139/ssrn.163151.

[4] B. Siliverstovs, D. van Dijk, Forecasting Industrial Production with Linear, Nonlinear, and Structural Change Models, Econom. Inst. Rep. EI 2003-16. (2002).

[5] M.P. Clements, D.F. Hendry, Forecasting Economic Time Series, Cambridge University Press (1998).

[6] B.E. Hansen, The New Econometrics of Structural Change, J. Econ. Perspect. 15 (2001) 117–128.

[7] A. Yin, Out-of-Sample Forecast Model Averaging with Parameter Instability, (2014) 1–40.

[8] G. Altansukh, & D. R. Osborn. Using structural break inference for forecasting time series. Empirical Economics, 63(2022), 1–41. https://doi.org/10.1007/s00181-021-02137-w.

[9] E. Bouri, A. Dutta, T. Saeed. Forecasting ethanol price volatility under structural breaks. Biofuels, Bioproducts and Biorefining, 15(2021) 250–256. https://doi.org/10.1002/bbb.2158

[10] L. Bauwens, & , A. Dufays, B. DE BACKER. Estimating and forecasting structural breaks in financial time series. Center for Operations Research and Econometrics, (2011).

[11] R. Kaushik, S. Jain, S. Jain, & T. Dash. Performance evaluation of deep neural networks for forecasting time-series with multiple structural breaks and high volatility. CAAI Transactions on Intelligence Technology, 6(2021) 265–280. https://doi.org/10.1049/cit2.12002

[12] G. C. Chow, Tests of Equality Between Sets of Coefficients in Two Linear Regressions. Econometrica, 28(1960) 591-605. https://www.jstor.org/stable/1910133

[13] R. L.Brown, J. Durbin, & J. M. Evans, Techniques for Testing the Constancy of Regression Relationships Over Time. Journal of the Royal Statistical Society:, 37(1975) 149–163. https://doi.org/10.1111/j.2517-6161.1975.tb01532.x

[14] B.Y.J. Bai, P. Perron, Estimating and Testing Linear Models with Multiple Structural Changes Author (s): Jushan Bai and Pierre Perron Published by : The Econometric Society Stable URL : https://www.jstor.org/stable/2998540, 66 (1998) 47–78.

[15] P.R. Winters, Forecasting Sales by Exponentially Weighted Moving Averages INFORMS Stable URL : https://www.jstor.org/stable/2627346, 6 (1960) 324–342.

[16] C.C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, 20 (2004) 5–10. doi:10.1016/j.ijforecast.2003.09.015.

[17] R.J. Hyndman, G. Athanasopoulos, Forecasting : Principles and Practice, OTexts, Melbourne, (2018).

[18] R.J. Hyndman, A.B. Koehler, J.K. Ord, R.D. Snyder, Forecasting with Exponential Smoothing: the state space approach, Springer-Verlag, Berlin, (2008).

[19] G. Box, G. Jenkins, Time series analysis: forecasting and control. Holden-Day (1970).

[20] R.J. Hyndman, Y. Khandakar, Automatic Time Series Forecasting : the forecast Package for R, 27 (2008). doi:10.18637/jss.v000.i00.

[21] R.F. Engle, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. The Econometric Society, 50 (1982) 987–1007.

[22] T. Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics , 31 (1986) 307–327.

[23] R.J. Hyndman, New in forecast 4.0. (2012). https://robjhyndman.com/hyndsight/forecast4/

[24] A. Soy Temür and Ş. Yıldız, Comparison of Forecasting Performance of ARIMA LSTM and HYBRID Models for The Sales Volume Budget of a Manufacturing Enterprise, İstanbul Business Research, 50 (2021),15-4

[25] C. N. Babu, & B. E. Reddy, A Moving-Average Filter Based Hybrid ARIMA–ANN Model for Forecasting Time Series Data. Applied Soft Computing, 23(2014), 27-38. doi:https://doi.org/10.1016/j. asoc.2014.05.028