Time Series Versus Causal Forecasting: An Application of Artificial Neural Networks

Prithviraj Lakkakula

Department of Agribusiness and Applied Economics, North Dakota State University
811 2nd Ave N., Fargo, ND 58102, Tel.: +1 701-231-6642,
Email: prithviraj.lakkakula@ndsu.edu

Abstract. This paper uses artificial neural networks to evaluate the performance of soybean and high fructose corn syrup (HFCS) price forecasting using both causal and time series techniques. Mean absolute deviation, and mean square error (both in–sample and out–of–sample) are used to evaluate the predictive accuracy of causal and time series neural networks. Based on the out–of–sample forecast performance, causal neural networks performed better in predicting both soybean and HFCS prices. To check the robustness of the results, a turning point test and sensitivity analysis are conducted. A turning point test is performed to evaluate the technique that captures the most turning points. Turning point test results indicate that the time series forecasting approach captures the most turning points for both soybean and HFCS prices. Finally, a sensitivity analysis is performed to analyze the relative importance of explanatory variables on soybean and HFCS prices. Results of this article help to guide forecasting approach selection based on research objectives.

Key words: Neural Networks; price forecasting; Out–of–sample performance; turning points

1 Introduction

Forecasting is an important area of research in various fields including agriculture, and engineering. Better forecasts lead to better decision making. For example, a prevalent strategy in any industry is to plan future inventory levels to maximize expected returns. However, obtaining better forecasts is a difficult task. Forecasting is even more difficult in agriculture as it is subjected to unpredictable weather, which increases volatility in yield and price.

Expectations about commodity prices, which are driven by several factors including expected demand, and supply, inform producers about crop production decisions. Commodity price forecasts are also important globally as they influence farm income. As commodity prices are frequently volatile, governments tend to support farmers to ease income volatility. For example, data compiled by the United States Department of Agriculture (USDA) indicate that net farm income projections decreased by 8.7% for the fourth consecutive year after a peak in 2013 [1]. Considering the significance of commodity prices and their impact on the macroeconomic activity of a country, their accurate prediction is important.

This article examines the performance of both causal and time series approaches of forecasting to determine which approach is better: time series or causal? To address this research question, neural networks are used for training monthly Illinois soybean freight on board (FOB) prices between 1986:2 and

2013:1, and quarterly HFCS-42 prices between 1995:1 and 2013:3. Later, the respective trained models are evaluated with monthly soybean prices between 2013:2 and 2016:1, and quarterly HFCS prices between 2013:4 and 2015:3.

Time series forecasting and causal forecasting are two main approaches of quantitative forecasting. To examine the relative popularity of the two techniques, a Google Scholar search was done using the two terms: "Time Series Forecasting" and "Causal Forecasting". For example, during 2010-2016, causal forecasting and time series forecasting counts are 300 and 16,300, respectively. The search results suggest that time series forecasting has been more popular among the research community compared to causal forecasting.

Time series forecasting uses the historical data of a variable to understand the autocorrelation and partial autocorrelation structures of the time series. This structural relationship between present and past data series is later used to develop and estimate the model. Causal forecasting, which is not widely used compared to time series forecasting, consists of two sets of variables: dependent and independent. Independent variables are chosen such that they explain a significant variation in the dependent variable. However, few methods such as vector autoregression seems to fit the definition of both the techniques.

1.1 Time Series and Causal Neural Network Studies

[2] were the first to model time series using neural networks. Subsequently, many studies have used neural networks for forecasting. [3] applied both autoregressive integrated moving average (ARIMA) and feed-forward neural networks to study monthly live cattle and wheat prices for the 1950–1990 period. For both price series, they found that neural networks performed better with lower mean square error as well as capturing maximum turning points compared to the ARIMA model.

[4] applied ARIMA (Autoregressive Integrated Moving Average), neural networks, and a hybrid model (using both the ARIMA and neural networks) to three different data sets, wherein the hybrid model performed well compared to the other two models based on the lowest mean square error and mean absolute deviation. Zhang's hybrid model consists of two steps [4]. First, estimate the ARIMA model to collect residuals of the model. Second, model these residuals in neural networks by incorporating lagged residuals as input variables.

[5] claims that data preprocessing is unnecessary when applying neural networks because they are considered as the universal approximators. However, subsequent studies claim that performing data preprocessing steps such as deseasonalizing and detrending the data yield better forecast accuracy [6],[7]. Specifically, [7] found that deseasonalizing and detrending the data before incorporating into the neural networks yield better results compared to raw data. Hence, in this article, three different data preprocessing steps including deaseasonalization, detrending, and Z-score normalization are performed.

[8] compared several machine learning techniques using monthly M3 time series competition data to evaluate a time series forecast. They found that data preprocessing has played an important role in the performance of the model.

Most previous studies have focused relatively more on time series forecasting compared to causal forecasting with neural networks. But, little research has focused on estimation and comparison of model performance, especially in the context of causal versus time series forecasting approaches and hence this study contributes to the literature. This comparison between causal and time series approaches is important to understand the type of approach to be used with specific types of data to make better and informed decisions. In this article, causal and time series approaches are used to predict both soybean prices and HFCS prices. Later, the performance of both causal and time series neural networks are compared through various forecast accuracy measures. Finally, turning point test results are discussed comparing causal and time series forecasting approaches.

2 Model Specification

2.1 Artificial Neural Networks

The artificial neural network contains several interconnected neurons (nodes) that exchange and process information through so-called weights. Weights are analogous to parameters, which are obtained during the process of training the network. A typical neural network consists of an input layer, hidden layer, and output layer. Each layer consists of nodes (equivalent to neurons). For example, the number of nodes in the input layer represent the number of input variables and the number of nodes in the output layer represent the number of output variables.

The hidden layer processes the information in the form of a signal received from the input layer. Signals are first processed by an *integration function* responsible for combining all incoming signals and second by an *activation function* responsible for transforming the output of the neuron. The weights or parameters in the network are assigned using a learning algorithm that minimizes an error function such as the mean square error. The inputs i into hidden neuron j are linearly combined to obtain the integration function

$$Z_j = \beta_{0j} + \sum_{i=1}^{3} \beta_{ij} X_i \tag{1}$$

where i, j, X_i, β_{0j}, and β_{ij} are the number of nodes in the input layer, number of nodes in the hidden layer, input variable i, the weight connecting the bias unit and hidden neuron j, and the weight connecting input neuron i and hidden neuron j. In the hidden layer, Z_j is then modified using a non-linear function (activation function) such as the sigmoid function $f(Z) = \frac{1}{1+exp(-Z)}$ to serve as an input for the next layer. Training the network is analogous to moving down an error surface, which takes place by adapting weights as per the learning algorithm rule [3]. The training process is an iterative process, which involves the following steps [9]:1) compute the predicted or forecast value with given inputs and estimated weights, 2) calculate the error sum of the squares, and 3) all weights are updated given the rule of the learning algorithm. The iterative process stops when the pre-specified criterion of error sum of squares is fulfilled if all absolute partial derivatives of the error function w.r.t weights are less than the threshold (i.e., $\frac{\partial E}{\partial \beta} < threshold$). The threshold levels are assigned based on how fast the model convergence occurs. In this estimation algorithm, 0.001 is the threshold level used for predicting HFCS prices and 0.01 is the threshold level used for predicting soybean prices. A resilient back propagation algorithm is used for training the network. All algorithms minimize the error function by adding a learning rate to the weights going in the opposite direction of the gradient.

The most important characteristic of the neural networks is the initialization of the weights. Unlike most traditional models, the initialization of the weights should be non-zero values. If the initial values of the weights are zero, then the final values of all the weights will be the same for all input variables. To prevent this issue, random values drawn from standard normal distribution are assigned as initial weights other than zero. In the output layer, there will be a choice of collecting information linearly or non–linearly. Usually, in the regression context, the most prevalent option is linear. In this article, the output layer consists of one node/variable, which is the explained variable–HFCS-42 spot price or soybean FOB price depending on the dataset used.

This article designed the following steps in the methodology: 1) data-splitting, 2) data preprocessing including deseasonalizing, detrending, normalization, 3) training the NN for both the causal and time series approaches, 4) predict HFCS and soybean prices using estimated modesl and respective test data, and 5) evaluate the forecasts using various forecast accuracy measures.

Two important decisions need to be made before training the neural networks: 1) the selection of the autoregressive terms of the variable that explains a significant variation in that variable, and 2) the number of hidden nodes to be specified in the hidden layer. [10] claim that the number of hidden nodes should be $\sqrt{m * n}$, where m is the number of input nodes and n is the number of output nodes. However, there is no “hard-and-fast rule” in deciding the number of hidden nodes, which are usually selected arbitrarily depending on the data. The model which gives the lowest out-of-sample MSE or out-of-sample MAD is selected as the best model. A precaution must be taken to not to overfit the model. Overfitting is a scenario where the out-of-sample MSE is very high and in-sample MSE is very low.

2.2 Neural Networks

Causal networks are a representation of a dependent variable as a function of independent variables. This article designed two causal networks: soybean price causal networks and HFCS price causal networks. In a time series model, a variable of interest is a dependent variable, which is specified as a function of its own lagged historical observations.

Soybean Price Network: Soybean is a highly traded agricultural commodity in the United States. Approximately 45% of the total US soybean production is exported (`https://www.mda.state.mn.us/food/business/~/media/Files/food/business/economics/exports-soybeans.ashx`). US soybean prices are affected by many factors including US soybean farm price, crude oil price, US Dollar exchange rate (captures the soybean trade impact). This article used the crude oil price, US dollar index (USDI) value, and one lagged US soybean farm price as the major factors explaining the variation in the soybean FOB price over the 1986–2016 period. During the process of model building, when Soybean exports were used as a proxy for production data to include in the model estimation, the results did not improve compared to the results when it was not included. The correlation results also indicate that the USDI captured most correlation patterns in Soybean FOB prices compared to the Soybean exports. Hence, the soybean model is represented as follows: $P_t = \alpha_0 + \alpha_1 * CrudeOilP_t + \alpha_2 * Soybean1LFP_t + \alpha_3 * USDI_t$ where P_t is the Illinois soybean FOB price (US\$/bushel) at time t, $CrudeOilP_t$ is the crude oil price (US \$/barrel) at time t, $Soybean1LFP_t$ is one lagged US soybean farm price (US \$/bushel) at time t, and $USDI$ is the US dollar index value at time t.

HFCS Price Network

HFCS is one of the major caloric sweeteners used in the United States. Currently, HFCS accounts for about 25–30% of the total caloric sweetener consumption in the United States (`http://www.ers.usda.gov/topics/crops/sugar-sweeteners/data.aspx`). Corn is the major raw material used in the production of HFCS. The quarterly corn price, total quantity of HFCS production, and total fructose exports are used as the major contributing factors that explain the significant variation in the HFCS-42 price. Causal neural networks of the HFCS price model has three nodes in the input layer, four nodes in the hidden layer, and one node in the output layer. Typically, in a regression context, there is only one output variable, except in classification models and/or binary response models. The presence of a hidden layer makes the neural network non-linear. The HFCS model is represented as follows:

$PH_t = \beta_0 + \beta_1 * CornP_t + \beta_2 * QtyHFCSP_t + \beta_3 * TFructoseExp_t$

where PH_t is the HFCS-42 spot price (US cents/pound) at time t, $CornP_t$ is the corn price (US \$/bushel), $QtyHFCSP_t$ is total quantity of HFCS produced in the United States at time t (short tons), and $TFructoseExp_t$ is the total US fructose exports at time t (metric tons). Neural networks can be written as [7]

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} x_{it} + \beta_{0j}\right) + \epsilon_t \tag{2}$$

where m, and n are the number of input nodes, and hidden nodes, respectively. While f is a sigmoid transfer function such as the logistic, $f(x) = \frac{1}{1+exp(-x)}$ or a $tanh$ function. $(\alpha_j, j = 0, 1, ..., n)$ is a vector of weights or parameters connecting the hidden and output nodes and $(\beta_{ij}, i = 0, 1, ..., m; j = 1, 2, ..., n)$ are the weights or parameters connecting the input and hidden nodes. Finally, α_0, β_{0j} are called bias terms, which are equivalent to the intercept (the bias terms always take a value of 1). The time series model is specified as: $Y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t$ where Y_t is the dependent variable (also the interested variable for forecast), y_{t-i} is the lagged y_t variables, ε_t is the residual at time t, and $i = 1, ...p$ is the number of lags.

Two features need to be determined in the time series model before incorporating into the neural networks: 1) the stationarity of the raw data, 2) the appropriate number of lags in the model. There are four different criteria to determine the appropriate number of lags to be used for model estimation. They are based on 1) the statistical significance test, 2) the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), 3) adding lags of the explained variable as the independent variable until the elimination of the serial correlation in the residuals, and 4) partial autocorrelation functions [11]. This article primarily used the third and fourth criteria to determine the appropriate number of lags to be included in the neural networks. Specifically, a residual analysis is performed to check for any significant information left out in the residuals.

One of the concerns of neural networks is that it is susceptible to a local minimum trap that can be mitigated by training the model multiple times while simultaneously experimenting with different hidden nodes with different starting values. The best model was selected based on the lowest test MSE or the lowest MAD.

This article employs the following forecast accuracy measures in evaluating both models: 1) In-sample mean square error: $IMSE = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - Y_t^*\right)^2$, 2) out-of-sample mean square error: $OMSE = \frac{1}{s}\sum_{t=1}^{s}\left(Y_t - Y_t^*\right)^2$, 3) In-sample mean absolute deviation: $IMAD = \frac{1}{T}\sum_{t=1}^{T}|Y_t - Y_t^*|$, and 4) Out-of-sample mean absolute deviation: $OMAD = \frac{1}{s}\sum_{t=1}^{s}|Y_t - Y_t^*|$. where Y_t is the actual value at time t, Y_t^* is the forecast value at time t, and T in IMSE and IMAD represents a total of 75 in-sample observations and 324 in-sample observations in the HFCS price model and the soybean price model, respectively. Also, s in OMSE and OMAB represents a total of 8 out-of-sample observations and 36 out-of-sample observations in the HFCS price model and the soybean price model, respectively.

2.3 Pesaran and Timmermann's (2009) Test for Turning Point Evaluation

Several studies have been proposed to study the directional accuracy of forecasts [12],[13],[14]. However, all the above tests are based on the assumption that null distributions are independently and identically distributed ($i.i.d$). A test proposed by [15](henceforth, PT Test) is used because it addresses the serial correlation between forecasts and actual binary sequences. The PT test evaluates the turning points comparing both the actual and predicted binary sequences of both the soybean price model and HFCS price model. Specifically, the turning points are evaluated separately for the training data, test data and whole (all sample) data.

The PT test tests the dependence of two binary sequences. The binary sequences are obtained after performing a forecast analysis. Specifically, X_t is defined as a binary sequence for the forecast variable that takes a value of 1 if ΔX_t is greater than 1, and zero otherwise. That is, if the change in the value of the forecast in time t and in time $t-1$ is greater than one, then X_t takes a value of 1, and zero otherwise. Similarly, Y_t is the binary sequence of the actual series that takes a value of 1 if ΔY_t is greater than 1, and zero otherwise. This means that for a time period, if X_t, and Y_t take the same value—either zero or one—then the forecast and actual time series moves in the same direction. The null hypothesis is that the direction of change in a forecast and the actual binary series is independent [15]. [15] proposed two versions of tests: the trace canonical correlation test and the maximum canonical correlation test. However, in the case of a two-dimensional two-way table (such as in this case), these two tests are similar. Due to space considerations the detailed PT test is not presented here. For more details on PT test see [15], [16], and [17].

3 Data

Data for the quarterly US HFCS–42 spot price (cents/pound), US total quantity of HFCS production ('000 short tons, dry weight), total US fructose exports (metric tons, dry weight basis), and US yellow corn price ($/bushel) are collected from the United States Department of Agriculture (USDA) sugar and sweetener yearbook tables and USDA feed grain tables for the 1995–2015 period (`http://www.ers.usda.gov/topics/crops/sugar-sweeteners/data.aspx`). The entire data are split into two groups: Training set (1995:Q1 to 2013:Q3) and Test set (2015:Q1 to 2015:Q3). Therefore, the training set consists of 75 observations while the test set consists of 8 observations accounting for nearly 10% of the sample size.

Similarly, data are collected for the monthly Illinois soybean FOB price ($/bushel), crude oil price (US$/barrel), a lagged US soybean farm price ($/bushel), and US dollar index value (no units) from USDA oilseeds crop outlook database, and World Bank for the 1986 (February)– 2016 (January) period. The soybean price model consists of 324 (1986:2 to 2013:1) observations in the training set and 36 (2013:2 to 2016:1) observations in the test set, which also accounts for about 10% of the sample size (`http://www.ers.usda.gov/data-products/oil-crops-yearbook.aspx`). The US dollar index data are collected from `http://www.investing.com/quotes/us-dollar-index-historical-data`.

It is important to note that the two datasets chosen for the analysis are distinct from each other in the sense that the HFCS data are not volatile and there is no continuous movement in the HFCS prices for the period considered. In contrast to HFCS prices, the soybean price data are more volatile and there is a continuous movement in the prices for the period considered. Representing two different types of data for the analysis ensure the generalization of the results to other datasets based on their characteristics.

Data preprocessing was performed to create more uniform data and to avoid computation problems. These data preprocessing steps included deseasonalization, detrending, and Z–score normalization. A versatile and robust method of decomposing time series for deseasonalization and detrending, called as "Seasonal and Trend decomposition using Loess" (STL) was used in the article [18].

Finally, Z–score normalization is performed to both the training and test data sets of the deseasonalized and detrended data by using the following formulas: For training data: $X_{trainnew} = \frac{X_{trainold} - mean_{train}}{std_{train}}$, and for test data: $X_{testnew} = \frac{X_{testold} - mean_{train}}{std_{train}}$. Where $X_{trainnew}$, $X_{testnew}$, $X_{trainold}$, $X_{testold}$, $mean_{train}$, and std_{train} are new training data, new test data, old training data, old test data, mean of training data, and standard deviation of training data, respectively, for both soybean prices and HFCS prices.

After convergence of the neural network to the given threshold value, the predicted values (also called in-sample forecasts) are computed. The model parameters and the out-of-sample independent test data are then used to predict the out-of-sample soybean prices and HFCS prices. Both the in-sample and out-of-sample predictions are then rescaled by seasonalizing and trending with the respective time periods. These rescaled predictions and the original prices of both the soybean prices and HFCS prices are finally used to compute their respective predictive accuracy measures such as the MSE and the MAD.

4 Results and Discussion

This section discusses the performance of both the causal and time series models and their ability to forecast soybean and HFCS prices. First, a method is chosen to decide the appropriate number of lags to be used in the model estimation of time series neural networks. Second, the selection of an appropriate number of hidden neurons employed in the neural networks is discussed. Third, forecast accuracy and turning point evaluation results are presented and discussed. Additionally, diagnostic tests are performed on the residuals of both the time series and causal neural networks of HFCS and soybean prices. Finally, the results are discussed in the broader context.

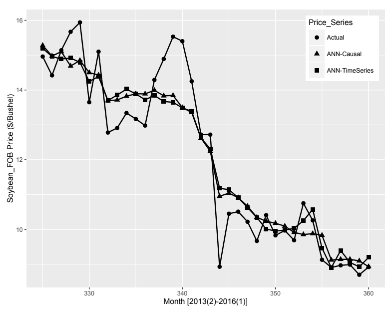
For time series neural networks, the unit root test was performed to ensure the stationarity of the data. This resulted in one differenced soybean prices and two differenced HFCS prices being stationary. Based on the partial autocorrelation functions (PACF), it was determined that five autoregressive terms (1st, 2nd, 3rd, 9th, and 23rd lags) of soybean prices and three autoregressive terms (1st, 2nd, and 3rd lags) of HFCS prices have explanatory power to predict Illinois soybean FOB prices and HFCS-42 spot prices, respectively. Therefore, five lagged explanatory variables were used for predicting soybean prices and three lagged explanatory variables were used for predicting HFCS prices in the respective time series neural networks.

For selecting the number of hidden neurons, the article explored several feed-forward networks with varying hidden neurons and later the best network with the lowest out-of-sample forecast accuracy measures was retained. More specifically, the article used hidden neurons ranging from one to ten consecutively and later chose the appropriate number of hidden neurons based on the best performed network. For soybean neural networks, the $3\times2\times1$ network was the best for causal while the $5\times7\times1$ network was the best for time series. Similarly, for HFCS neural networks, the $3\times1\times1$ network was the best for causal while the $3\times7\times1$ network was the best for time series. In the network, the first number indicates the number of input (explanatory) variables, second number indicates the number of hidden neurons, and the final number indicates the output (explained) variable.

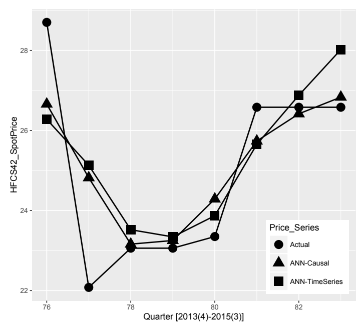
Table 1 shows the forecast accuracy results of both the causal and time series networks in terms of mean square error (MSE), and mean absolute deviation (MAD) for both in-sample and out-of-sample. Results of out-of-sample accuracy measures in table 1 show that causal forecasting performed better compared to time series forecasting in both soybean and HFCS price networks. This is a very important result given the fact that the time series forecasting technique has been more popular compared to causal forecasting. Also, the time series forecasting approach for both soybean and HFCS price networks has consistently performed well in the in-sample, but failed to generalize to the out-of-sample observations. Figure 1 show the out-of-sample fit for soybean and HFCS prices using both causal and time series networks. Specifically, the out-of-sample fit for HFCS prices indicates that the causal approach has performed better compared to the time series approach based on various forecast accuracy criteria.

Table 1: Forecast accuracy results of neural networks

Model Type	In-sample MSE	In-sample MAD	Out-of-sample MSE	Out-of-sample MAD
Soybean Model				
ANN–Causal	0.29	0.36	0.63	0.63
ANN–Time Series	0.23	0.33	0.66	0.64
HFCS Model				
ANN–Causal	0.26	0.38	1.67	0.91
ANN–Time Series	0.02	0.10	2.34	1.18



(a) Soybean price



(b) HFCS price

Fig. 1: Actual vs predicted soybean prices of both causal and time series neural networks (out-of-sample)

Table 2 shows the PT test results for the turning point evaluation [15]. Turning point evaluation results show alternative findings to the forecast accuracy results. The null hypothesis is rejected in all cases at least at the 1% significance level except in the case of the in-sample causal model of soybean prices, and the out-of-sample causal and time series models of HFCS prices. A rejection of the null hypothesis of the PT test implies that a forecast is a useful predictor of the actual change in the variable. Specifically, in the case of the out-of-sample soybean model, even though both the causal and time series approaches of forecasting are strong predictors of the change in direction of the actual variable, the time series forecasting approach predicted more turning points compared to the causal approach. The turning point evaluation results of HFCS prices show that in-sample and all sample results strongly reject the null hypothesis at the 1% level, while the time series approach captured the change in the direction of the actual variable more accurately compared to the causal approach. As a whole, the turning point evaluation results indicate that the time series approach is better suited when the objective is to maximize capturing turning points in the interested time series.

Table 2: Test results of turning point evaluation

	Pesaran and Timmerman (2009) Test Statistic		
Model Type	In-sample	Out-of-sample	All sample
Soybean Model			
ANN–Causal	2.29	6.21^{***}	5.41^{***}
ANN–Time Series	10.74^{***}	7.07^{***}	16.22^{***}
HFCS Model			
ANN–Causal	9.67^{***}	0.44	8.36^{***}
ANN–Time Series	16.30^{***}	1.33	14.74^{***}

Notes: ***p<0.01.

Additionally, two diagnostic tests were performed in order to validate the specification of both the causal and the time series neural networks. The Jarque-Bera (JB) test is used to test for the normality of the residuals , and the Augmented Dickey-Fuller (ADF) test is used to test for the stationarity of the residuals [19],[20]. The null hypothesis of the Jarque-Bera test is the normality of the residuals and it is rejected if the JB test statistic is larger than the critical value of the (χ^2) distribution. For both the soybean and HFCS models, the JB test statistic is less than the 5% critical value so the null hypothesis is not rejected. The null hypothesis of the ADF test is that the residuals are non-stationary meaning that there is a significant amount of information that is left in the residuals. All tests reject the null hypothesis and conclude that the residuals of all the models are stationary at the 5% level. Based on the above diagnostic tests, it is reasonably concluded that both the causal and time series networks of soybean and HFCS prices are correctly specified.

In summary, results of the article indicate that the use forecasting approach depends on the objectives of the research. If the objective is forecast accuracy, then based on the results, it is recommended to employ causal forecasting technique. On the other hand, if the objective is to capture maximum number of turning points, then time series forecasting technique is appropriate.

Finally, a sensitivity analysis was performed in this study. The primary goal of a sensitivity analysis is to evaluate the relative importance of explanatory variables based on their effect on the output variable. Two methods were chosen for sensitivity analysis including Garson's algorithm [21], and Lek's profile method [22],[23]. Garson's algorithm uses the connection weights between each of the input variables to quantify the effect of inputs on the output. A higher connection weight value between an input variable and the output variable indicates that a specific input variable has a higher impact on the output variable. On the other hand, Lek's profile method analyzes the effect of a change in an input variable on an output variable maintaining all other input variables at a certain value. Specifically, it constructs a fictitious matrix pertaining to the range of all input variables in which only one input variable is varying at a time and all other input variables are fixed at certain values such as the minimum, first quartile, median, third quartile, and maximum. Therefore, the effect of each input variable on the output variable generates a plot of five profiles. For additional details of both Garson's algorithm, and Lek's Profile method, refer to [23].

Garson's algorithm results show that the most important variable for determining the soybean price is US Dollar Index, while the least important (of three input variables) is crude oil price. Lek's profile results indicate that the soybean price increases with increase in the crude oil price when all other input variables are constant at the respective quartiles. An increase in the one lagged soybean farm price has positive or increased expectation on the soybean fob price except for minimum. Finally, a higher US Dollar Index has a negative effect on the soybean price for all profile groups except for minimum.

5 Conclusions

This paper demonstrates the forecast performance of two approaches: causal and time series using artificial neural networks, which was neglected in the literature. Both neural networks are used for predicting Illinois soybean FOB prices and HFCS-42 spot prices. Forecast accuracy measures such as mean square error and mean absolute deviation in the out-of-sample reveal that the causal approach performed better in predicting both soybean and HFCS prices compared to time series networks.

Turning point evaluation results suggest that the time series approach captured most of the turning points in both the in-sample, and out-of-sample soybean and HFCS prices. This implies that time series approaches are superior to causal approaches when considering turning point analysis. Even though time series approach seem to be more popular, results of this article indicate that the use of forecasting approach depends on the objective of the researcher. For example, if the researcher is interested in capturing maximum turning points, then the results of the article suggest that the time series approach is more appropriate. On the other hand, if the researcher is interested in the accuracy of the forecast, then the researcher needs to employ the causal forecasting approach.

Sensitivity results of soybean causal model suggest that US Dollar Index is the most important variable that affects the soybean FOB price, while results of HFCS causal model suggest that corn price is significant variable in determining the HFCS-42 price.

References

1. Highlights From the February 2017 Farm Income Forecast, `https://www.ers.usda.gov/topics/farm-economy/farm-sector-income-finances/highlights-from-the-farm-income-forecast/`.
2. Lapedes, Alan, and Robert Farber. Nonlinear signal processing using neural networks: Prediction and system modelling. No. LA-UR-87-2662; CONF-8706130-4. 1987.
3. Kohzadi, Nowrouz, Milton S. Boyd, Bahman Kermanshahi, and Iebeling Kaastra. A comparison of artificial neural network and time series models for forecasting commodity prices. Neurocomputing 10, no. 2 (1996): 169-181.
4. Zhang, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50 (2003): 159-175.

5. Gorr, Wilpen L. Research prospective on neural network forecasting. International Journal of Forecasting 10, no. 1 (1994): 1-4.
6. Nelson, Michael, Tim Hill, William Remus, and Marcus O'Connor. Time series forecasting using neural networks: Should the data be deseasonalized first?. Journal of forecasting 18, no. 5 (1999): 359-367.
7. Zhang, G. Peter, and Min Qi. Neural network forecasting for seasonal and trend time series. European journal of operational research 160, no. 2 (2005): 501-514.
8. Ahmed, Nesreen K., Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews 29, no. 5-6 (2010): 594-621.
9. Gnther, Frauke, and Stefan Fritsch. neuralnet: Training of neural networks. The R journal 2, no. 1 (2010): 30-38.
10. Kaastra, Iebeling, and Milton Boyd. Designing a neural network for forecasting financial and economic time series. Neurocomputing 10, no. 3 (1996): 215-236.
11. Hill, R. Carter, William E. Griffiths, and Guay C. Lim. Principles of econometrics. Vol. 5. Hoboken, NJ: Wiley, 2008.
12. Henriksson, Roy D., and Robert C. Merton. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. Journal of business (1981): 513-533.
13. Cumby, Robert E., and David M. Modest. Testing for market timing ability: a framework for forecast evaluation. Journal of Financial Economics 19, no. 1 (1987): 169-189.
14. Pesaran, M. Hashem, and Allan Timmermann. A simple nonparametric test of predictive performance. Journal of Business & Economic Statistics 10, no. 4 (1992): 461-465.
15. Pesaran, M. H. and Timmermann, A: Testing dependence among serially correlated multicategory variables. Journal of the American Statistical Association, 104(485), 325–337 (2009)
16. Chou, Cheng, and Chia-Shang J. Chu. Market timing: Recent development and a new test. Economics Letters 111, no. 2 (2011): 105-109.
17. Tsuchiya, Yoichi. Are government and IMF forecasts useful? An application of a new market-timing test. Economics Letters 118, no. 1 (2013): 118-120.
18. Cleveland, Robert B., William S. Cleveland, and Irma Terpenning. STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6, no. 1 (1990): 3.
19. Jarque, Carlos M., and Anil K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters 6, no. 3 (1980): 255-259.
20. Said, Said E., and David A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, no. 3 (1984): 599-607.
21. Garson, David G. Interpreting neural network connection weights. (1991): 47-51.
22. Lek, S., A. Belaud, I. Dimopoulos, J. Lauga, and J. Moreau. Improved estimation, using neural networks, of the food consumption of fish populations. Oceanographic Literature Review 9, no. 43 (1996): 929.
23. Gevrey, Muriel, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological modelling 160, no. 3 (2003): 249-264.