Research Article

Corn Price Modeling and Forecasting Using Box-Jenkins Model

Paibool Kitworawut and Vichai Rungreunganun*
Department of Industrial Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

* Corresponding author. E-mail: vichair@kmutnb.ac.th DOI: 10.14416/j.asep.2019.02.007
Received: 30 October 2017; Revised 6 January 2018; Accepted: 15 January 2018; Published online: 21 February 2019

Abstract
Corn harvesting is one of the most complicated problems which farmers need information to make decision prior farming. Corn price is the main factor for farming and there are many factors that affect the corn price vice versa. Knowledge of the factors affecting the corn price and the ability to forecast the corn price in advance would benefit farmers in the context of harvesting. The factors that affect the corn price in Thailand include chicken export rate, corn import rate, weather, soybean price, corn production, stock-to-use, season and planting area. The Cause Tree diagram has been constructed to demonstrate the linkage of such factors and all related data have been collected and analysed by using SPSS software. The Box-Jenkins model has been implemented to establish a time series forecasting model. And performance comparisons among the ARIMA model with Holt-Winters multiplicative seasonal model and Holt-Winters additive seasonal model methods. The results of this research indicated that the corn price can be forecast by using its two lag data with current period soybean price data. The resulting forecasting equation with the ARIMA model generate the lowest errors with Root Mean Square Error (RMSE) at 0.8678, Mean Absolute Percent Error (MAPE) at 12.1009 and Mean Absolute Error (MAE) of 4.7592.

Keywords: The cause tree diagram, SPSS, Box-Jenkins model, Holt-Winters Multiplicative Seasonal model, Holt-Winters Additive Seasonal model

1 Introduction

Corn is one of the most important agricultural products and is critical to both humans and animals. It can be used in variety applications, e.g., pharmacological activities [1]. Pimentel and Patzek [2] mentioned the use of corn as raw material for ethanol production, starch, food for human, animal feeds. Corn may even be used in high fructose syrup production as proposed by Parker *et al*. [3].

In the market in Thailand, corn is used as the main feed for many kinds of livestock, e.g., chickens, pigs, etc. and constitutes more than 40% of pigs and more than 30% of chicken feed ingredients. Although, the demand for corn is quite high but its price fluctuates greatly and knowledge of the factors that affect corn price would beneficial to farmers in the context of corn harvesting. There has been plenty of research that has taken places regarding the factors that affect the corn price but only in respect of certain dimensions. Whittaker [4] studied only the effects of planted acreage of corn, corn yield, weather and ethanol production in determining the corn price. Wescott and Hoffman [5] have established a corn price forecasting model and found that the stock-to-use of corn also affects the corn price. These authors also mentioned that corn and soybeans compete with each other in term of farmers production decisions. Government programs have also

Please cite this article as: P. Kitworawut and V. Rungreunganun, "Corn price modeling and forecasting using Box-Jenkins model," *Applied Science and Engineering Progress*, vol. 12, no. 4, pp. 264–276, Oct.–Dec. 2019.

been important in influencing farm-level prices of corn and wheat. Condon *et al*. [6] mentioned that each billion gallon expansion in ethanol production yields a 2–3% increase in corn prices on average.

It is important not only to know the factors that affect the corn price but also the relationships among the factors are very important. Knowing these relationships would bring more insight and understanding to the system as a whole. System thinking facilitates the viewing of a system from more than one perspective and the cause tree diagram could then be developed to demonstrate the linkage among each potential factor in a graphical manner.

Collected data from related factors would yield a time series pattern. These time series data might compose variety of components, e.g., trend, cyclical, seasonal or even irregular components might be combined in different ways. The Autoregressive Integrated Moving Average or ARIMA (p,d,q) model was proposed by Box and Jenkins [7] and represents a methodology to decompose such components.

The purposes of this research are to model the relationships among each of the factors affecting the corn price in order to establish a corn price forecasting model. The forecasting model is then tested with actual data and compared with the forecasting performance of other methods.

2 Methods

The research was designed to take place in 2 phases with a total of 8 steps as given below. The first phase sought to determine the independent variables by using system thinking. And the second phase established the forecasting model by using the ARIMA (Box-Jenkins) model.

1. System thinking.
    1.1 Develop the cause tree diagram.
    1.2 List all independent variables.
2. ARIMA
    2.1 Data preparation.
    2.2 Model selection.
    2.3 Parameter estimation.
    2.4 Model checking.
    2.5 Forecasting.
    2.6 Compare with other methods.

Corn prices data have been collected with 240 data points gathered during January 1997 to December 2016 from Office of Agricultural Economics source. These data were separated into 2 parts, the first part comprises 228 data points during January 1997 to December 2015 and were used to generate the forecasting model. The second part comprises 12 data points collected during January 2016 to December 2016. These latter data were used to test the forecasting model and to compare the results with other forecasting methods.

2.1 *System thinking*

System thinking has its foundation in the field of system dynamics, as pioneered by Forrester [8] at the Massachusetts Institute of Technology. It models the relationships between elements in a system and how these relationships influence the behavior of the system over time. The objective of systems thinking modeling is to improve our understanding of the ways in which an organization's performance is related to its internal structure and operating policies, including those of customers, competitors and suppliers and then to use that understanding to design high leverage policies for success [8].

System thinking models have been implemented to capture the linkage of factors in models for discrete time series and dynamic systems. The models may be used to obtain optimal forecasts and optimal control action. Such models also form a connection between structure and decisions that generate system behavior. The cause tree diagram is an important tool for representing the linkage of factors of in the system and has long been used in academic work and is also increasingly common in business [8].

2.2 *ARIMA*

A time series is a collection of data recorded over a period of time be it weekly, monthly, quarterly or yearly. Time series forecasting is a statistical process which uses historical data to describe a pattern of correlations among variables. The model can be generated and then used to forecast future data. The analysis of history represents an interesting use of time series data and decisions and plans may be made based on the forecast results. In any time series data might comprise many types of variation, e.g., trends, cyclical features, seasonal or even irregular variations, which lead to inaccuracy of the forecasting result. The Box-

Jenkins methodology, known as the Autoregressive Integrated Moving Average (ARIMA), eliminates these variations and improves the accuracy of forecasting. The time series model using the Box-Jenkins approach was proposed by Box and Jenkins since 1970 [8]. The model is concerned with the building of frameworks to describe discrete time series and dynamic systems. The models may be used to obtain optimal forecasts and optimal control actions.

Box-Jenkins analysis refers to a systematic method of identifying, fitting, checking and using the ARIMA time series models. ARIMA (p,d,q) models are the extension of the autoregressive (AR) model that uses three components for modeling the serial correlation in the time series data. The first component is the AR term. The AR (p) model uses the p lags of the time series in the equation. The second component is the integration (d) order term. Each integration order corresponds to differencing the time series. I (d) means differencing the data d times. The third component is the Moving Average (MA) term. The MA (q) model uses the q lags of the forecast errors to improve the forecast.

The original Box-Jenkins modeling procedure involved an iterative three-stage process comprising model selection, parameter estimation and model checking [8]. Recent explanations of the process often add a preliminary stage of data preparation and the last stage of forecasting [9], as demonstrated in Figure 1.

The first stage, data preparation, involves transforming and differencing the data to produce a series compatible with the assumption of stationarity. The second step, model identification, finds the most satisfactory ARMA (p,q) model to represent the stationary data from the first stage by examining the autocorrelations function (ACF) and partial autocorrelation function (PACF) of the stationary series. The third stage, parameter estimation, finds the tentative values of the model coefficients which provide the best fit to the data. The fourth stage, model checking, involves testing the model. This stage passed the chosen value of ARIMA model and estimated its parameters. The adequacy of the model would be checked by analysing the value of its residuals. If the residuals are white noise, the model would be accepted; otherwise, the model is rejected and the process backward to start over stage 1 again. The last stage is forecasting. Once the model has been selected, the forecasting task would be accomplished.

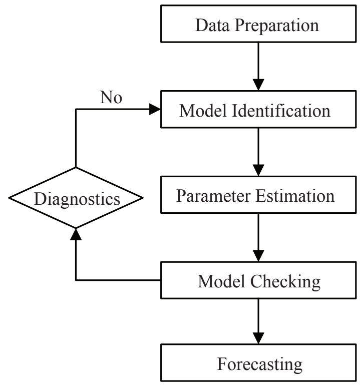

Figure 1: Box-Jenkins modelling procedure.

The ARIMA model is widely used in various applications, e.g., for short term forecasting of Hepatitis C Virus (HCV) seropositivity among volunteer blood donors in Karachi, Pakistan [10], for short-term market predictions, especially in the case of the high technology market which is characterized by a short life cycle due to rapid technology substitution [11], to forecast the major airline fatalities in the World using univariate time series models [12]. Weron [13] implemented ARIMA to forecast electricity price with various applications and the model has even been used for traffic condition prediction, proposed by Williams and Hoel [14].

Riansut and Thongrit [15] have constructed a forecasting model to predict the prices of field corn in Thailand during January 1997 to November 2015 by using the Box-Jenkins method. However, these authors used only previous values of the corn prices to construct the forecasting model. Kerdsomboon and Varaphakdi [16] have studied 4 agriculture products including rice, corn, green beans and soybean. The objectives were to construct appropriate forecasting models for plant acreage, products and prices. However, the forecasting models were generated separately in the absence of interrelation studies.

3 Results

3.1 *Cause tree diagram*

In regard to the collected potential factors proposed by Kitworawut and Rungreunganun [17] and to bring more insight understanding by demonstrating the

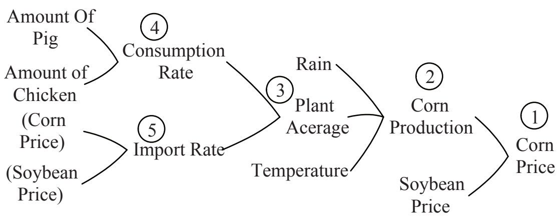

Figure 2: The cause tree diagram.

relationships among the variables a cause tree diagram may be constructed, as shown in Figure 2.

There are two types of variables, dependent and independent variables. From the cause tree diagram in Figure 2 the data can be classified into 5 groups with one additional group relating to the test of the direct effects from rain, plant acreage, temperature, consumption rate, import rate, amount of pigs and number of chickens on the corn price, as illustrated in Table 1. To test for any relationships among each group, dependent and independent variables in Table 1 are analysed by using cross-correlation function.

Table 1: Classification of dependent and independent variables from the causes tree diagram

Group	Dependent Variable	Independent Variable
1	Corn Price	Corn Production
		Soybean Price
2	Corn Production	Rain
		Plant Acreage
		Temperature
3	Plant Acreage	Consumption Rate
		Import Rate
4	Consumption Rate	Amount of Pig
		Amount of Chicken
5	Import Rate	Corn Price
		Soybean Price
6	Corn Price	Rain
		Plant Acreage
		Temperature
		Consumption Rate
		Amount of Pigs
		Amount of Chickens
		Import Rate

3.2 *ARIMA*

The corn price data have been collected with a total of 240 data points during January 1997 and December 2016 from Office of Agricultural Economics source. The data are separated into 2 parts. The first part comprises 228 data points gathered between January 1997 and December 2015. These are used to analyse the pattern of data. The second part, the data collected between January 2016 and December 2016, are used for corn price forecasting and comparing of the forecasting performance with other methods. This research has been analysed by using SPSS statistics software version 21.

In regards to Table 1, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are be used to determine the correlation between dependent variable data and lag time data itself. Then, the next step is to test the cross-correlation between the noise residuals from the dependent and independent variables for each group.

3.2.1 Corn price data with corn production and soybean price data

The corn price time series data have been plotted in Figure 3. This figure demonstrates that the data comprise trend variations. The ACF and PACF would be used to determine the correlation between the data and lag time itself. The ACF and PACF are illustrated in Figure 4.

From Figure 4, it is seen that ACF trends to decrease exponentially. The value of the PACF is significantly higher than zero at 2 lags. Hence, the appropriate time series model to forecast the corn price is the ARIMA (2,0,0). The results are demonstrated in Table 2.

Table 2: Model fit statistics

Model Description						
					Model Type	
Model ID cornprice Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.949	.949	.410	.4641	14.624	16	.552
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Cornprice-Model1 cornprice Constant			5.939	.798	7.446	.000
AR Lag 1			1.179	.065	18.076	.000
Lag 2			–.208	.065	–3.189	.002

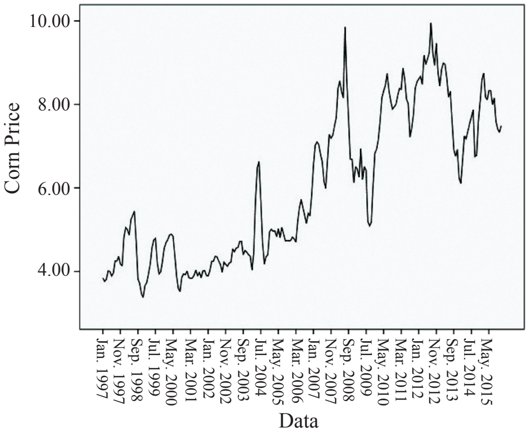

Figure 3: Graph of time series data of corn price.

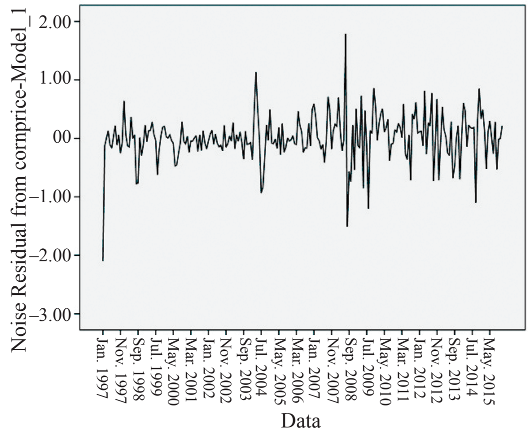

Figure 5: Graph of noise residual from corn price.

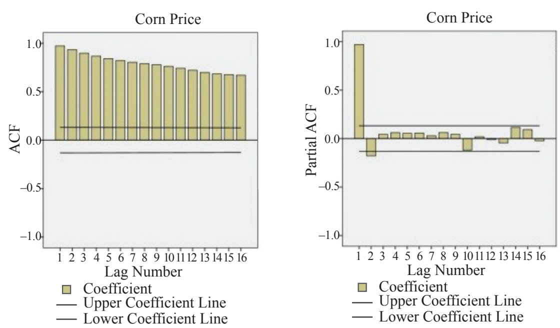

Figure 4: Graph of ACF and PACF of corn price data.

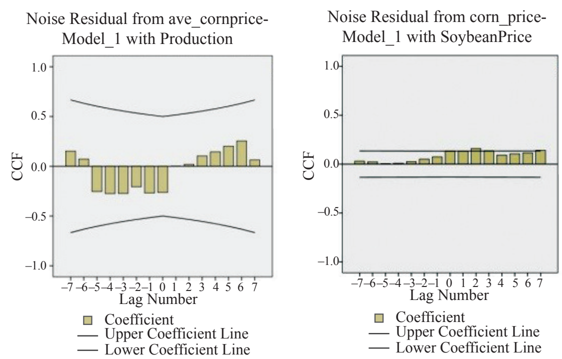

Figure 6: Graph of cross-correlation function between the noise residual from the corn price with the corn production and soybean price.

A plot of the noise residual from the corn price data is illustrated in Figure 5. It showed that the data was distributed around zero which means that the variance is constant. The results from Table 1 for the ARIMA (2,0,0) model indicate that the Ljung-Box Q statistics value is 0.626, which has no significant at value 0.05. This yields the conclusion that there is no autocorrelation of random errors for the ARIMA (2,0,0) model. The forecasting model for ARIMA (p,d,q) is demonstrated in Equation (1).

$$Y_t = (1-\Phi_1 - \ldots - \Phi_p)\mu + \Phi_1 Y_{t-1} + \ldots + \Phi_p Y_{t-p} + a_t \quad (1)$$

Hence, ARIMA (2,0,0) forecasting model for the corn price from the model parameters would be given by Equation (2).

$$Cornprice_t = 0.1739 + 1.19\ Cornprice_{t-1} - 0.221\ CornPrice_{t-2} \quad (2)$$

The r-square value from Equation (2) is 0.948 which means that this forecasting equation can explain the variation of the corn price at 94.8% with a Mean Absolute Percent Error (MAPE) at 4.645.

Next step is to check which factors affect the corn price by using cross-correlation function.

From Figure 6 it is found that only the soybean price has a correlation with the noise residual from the corn price. Hence, the analysis can be performed with the ARIMA (2,0,0) model with the dependent variable as corn price, and the independent variable as soybean price.

The equation for corn price forecast using the previous value of the corn price and the current price of soybean price is indicated in Equation (3).

$$Cornprice_t = 0.2098 + 1.161\ Cornpricet_{t-1} - 0.196\ Cornprice_{t-2} + 0.068\ Soybeanpricet \quad (3)$$

From Table 3 and Equation (3), it is possible to conclude that this forecasting equation can explain the variation of corn price at 95.1% with MAPE at 4.653.

Table 3: Model fit statistics

Model Description						
					Model Type	
Model ID corn price Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.951	.951	.402	4.653	16.461	16	.421
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Corn price Constant			5.099	.778	6.550	.000
AR Lag 1			1.161	.066	17.694	.000
Lag 2			–.196	.066	–2.984	.003
SoybeanPrice Numerator Lag 0			.068	.031	2.154	.032

3.2.2 Corn production data with rain, plant acreage and temperature data

Consider the corn production time series data in Figure 7. This figure indicates that the production increased until 2013 and then turned down.

Considering ACF and PACF for corn production data in Figure 8. The graph of ACF trends to decrease in damped sine wave pattern and the graph of PACF is indicated that the value of PACF after lag 2 is zero. Hence, the appropriate time series model to forecast corn production is AR (2). The analysis results are demonstrated in Table 4.

Table 4: Model fit statistics

Model Description						
					Model Type	
Model ID corn production Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.801	.801	209527.013	2.924	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Corn Production Constant			4570864.937	107826.590	42.391	.000
AR Lag 1			1.581	.124	12.705	.000
Lag 2			–.891	.111	-8.044	.000

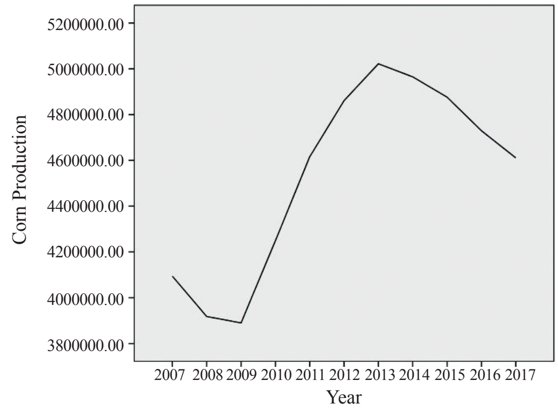

Figure 7: Graph of time series for corn production data.

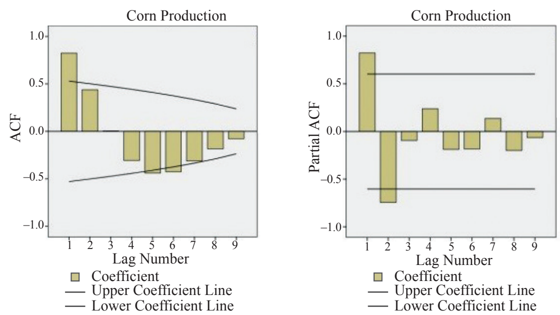

Figure 8: Graph of ACF and PACF for corn production data.

From the forecasting model for ARIMA (p,d,q) in Equation (1). Hence, ARIMA (2,0,0) forecasting model for corn production from model parameters would be in Equation (4).

$$CornProdcution_t = 1{,}404{,}338.581 + 1.581\ CornProduction_{t-1}\ 0.891 - CornProduction_{t-2} \quad (4)$$

It can be concluded that the Equation (5) can forecast corn production with an r-square of 80.1% and MAPE at 2.924.

The cross-correlation function in Figure 9 indicates that plant acreage has an effect on corn production.

From the forecasting model for ARIMA (p,d,q) in Equation (1) and data from Table5. They can be determined that the ARIMA (2,0,0) forecasting model for corn production using plant acreage would be described by Equation (5).

$$CornProdcutiont = 8{,}253{,}886.76 - 0.317\ CornProduction_{t-1} - 0.505\ CornProduction_{t-2} + 1.046\ PlantAcreage_t \quad (5)$$

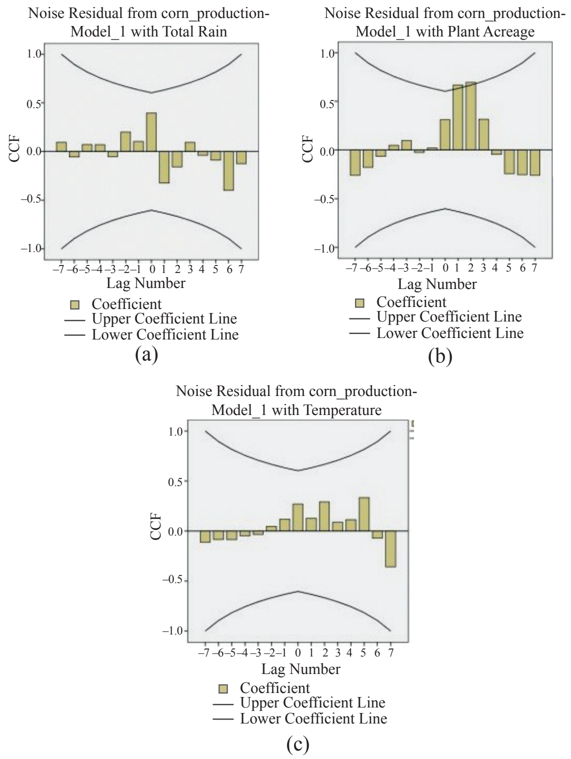

Figure 9: Graph of cross correlation function between noise residual from corn production with total rain (a), plant acreage (b) and temperature (c).

Table 5: Model fit statistics

Model Description						
					Model Type	
Model ID corn production Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics				**Ljung-Box Q(18)**		
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.922	.922	140110.983	1.881	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Corn Production Constant			–2830958.116	473022.241	–5.985	.001
AR Lag 1			–.317	.449	–.706	.503
Lag 2			–.505	.322	–1.569	.161
Plant Acreage Numerator Lag 0			1.046	.067	15.639	.000

It can be concluded that Equation (7) can forecast corn production using plant acreage with an r-square of 92.2% and MAPE at 1.881.

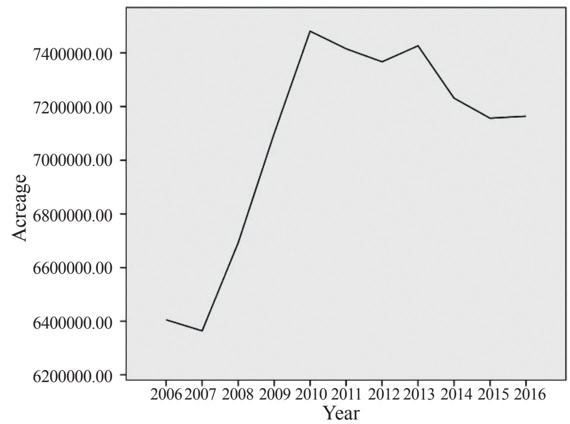

Figure 10: Graph of time series for plant acreage data.

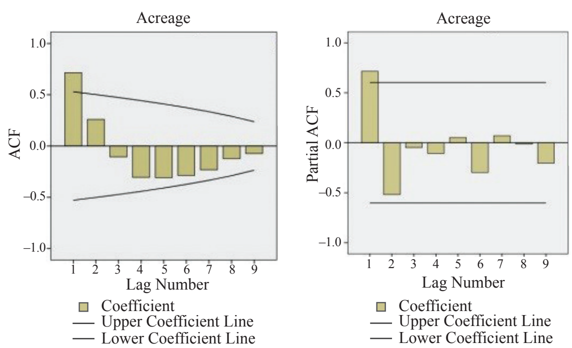

Figure 11: Graph of ACF and PACF for plant acreage.

3.2.3 Plant acreage data with consumption rate and import rate data

Consider the plant acreage time series data in Figure 10. This figure indicates that the plant acreage dramatically increased from 2007 until 2010 and then gradually decreased.

Considering the ACF and PACF for the plant acreage data in Figure 11, the appropriate time series model to forecast plant acreage is the ARIMA (2,0,0) model. The results are demonstrated in Table 6.

The forecasting model for the plant acreage data from the model parameters would be as shown in Equation (6).

$$PlantAcreage_t = 2{,}447{,}241.96 + 1.395\ PlantAcreage_{t-1} - 0.741\ PlantAcreaget_{-2} \quad (6)$$

From this equation, r-square value is 0.596. It means that this equation can explain the variation of plant acreage

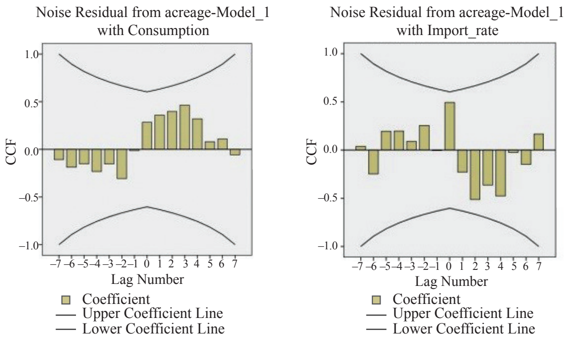

Figure 12: Graph of cross correlation function between noise residual from plant acreage with consumption rate and import rate.

Table 6: Model fit statistics

Model Description						
					Model Type	
Model ID acreage Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.596	.596	286928.272	2.551	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Acreage Constant			7096425.223	154728.341	45.864	.000
AR Lag 1			1.395	.222	6.294	.000
Lag 2			–.741	.201	–3.695	.006

at 59.6% with mean absolute percent error at 2.551.

To test for correlation function between noise residual from plant acreage with consumption rate and import rate.

Figure 12 demonstrates that neither import rate nor consumption rate have an effect on plant acreage.

3.2.4 Consumption rate data with numbers of pig and numbers of chicken data

Consider the consumption time series data in Figure 13, which indicates that the corn consumption rate is increasing continuously.

Considering the ACF and PACF for the consumption rate in Figure 14, the graph of ACF trends to decrease in a damped sine wave pattern. The appropriate time series model to forecast plant acreage is the ARIMA (2,0,0) model. The results are demonstrated in Table 7.

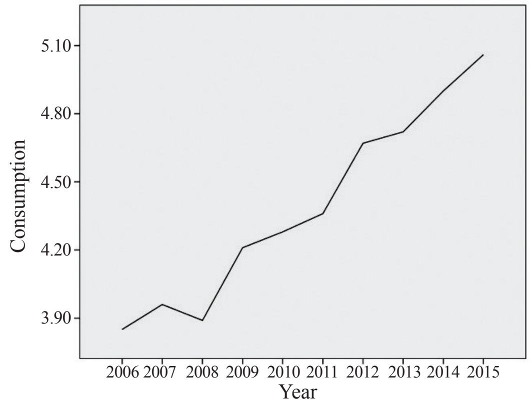

Figure 13: Graph of corn consumption rate annually.

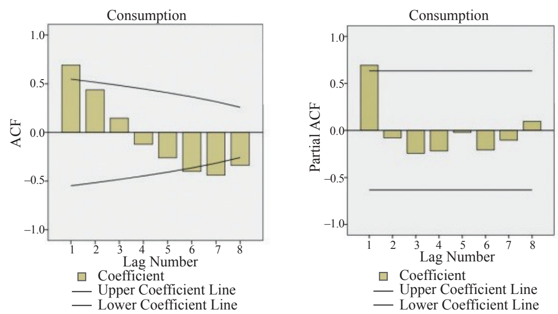

Figure 14: Graph of ACF and PACF for corn consumption rate data.

Table 7: Model fit statistics

Model Description						
					Model Type	
Model ID consumption Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.642	.642	.292	4.291	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Consumption Constant			4.436	.968	4.582	.003
AR Lag 1			1.215	.395	3.077	.018
Lag 2			–.287	.480	–.598	.569

The forecasting model for consumption rate from the model parameters are as shown in Equation (7).

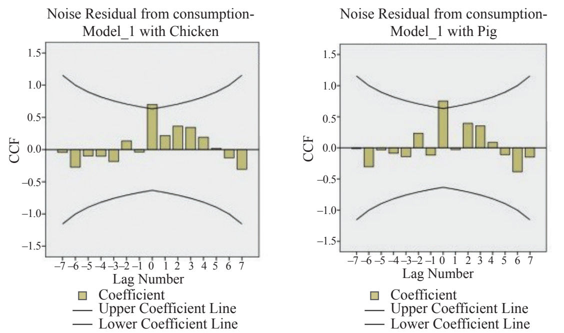

Figure 15: Graph of cross correlation function between noise residual from consumption rate with quantity of pigs and chickens.

$$ConsumptionR_t = 0.316 + 1.215\ ConsumptionR_{t-1} - 0.287\ ConsumptionR_{t-2} \quad (7)$$

It could be concluded that Equation (7) can forecast the consumption rate by using its previous value with an r-square for 0.642 and MAPE of 4.291.

To test for the correlation function between the noise residual from consumption rate with the number of pigs and chickens.

Figure 15 indicates that the quantities of both pigs and chickens with the noise residual of consumption rate. Hence, the analysis can be performed with the ARIMA (2,0,0) model for dependent variable consumption rate, and with the independent variable as the quantity of pigs and chickens. The analysis result are demonstrated in Table 8.

Table 8: Model fit statistics

Model Description						
					Model Type	
Model ID consumption Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.945	.945	.135	2.042	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Consumption Constant			3.544	.595	5.960	.002
AR Lag 1			.036	.538	.066	.950
Lag 2			–.235	.465	–.505	.635
Pig Numerator Lag 0			–1.692E-7	1.039E-7	–1.627	.165
Chicken Numerator Lag 0			7.613E-9	1.355E-9	5.620	.002

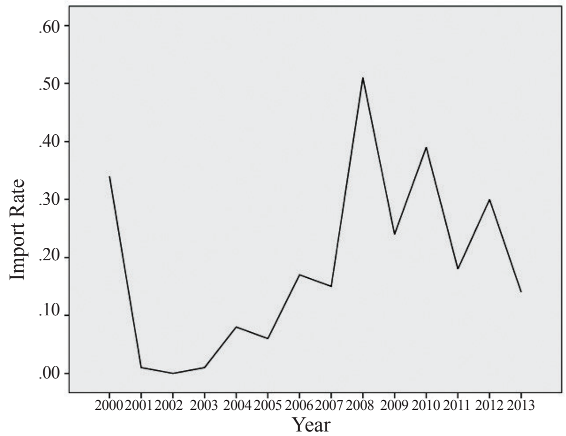

Figure 16: Graph of time series for import rate data.

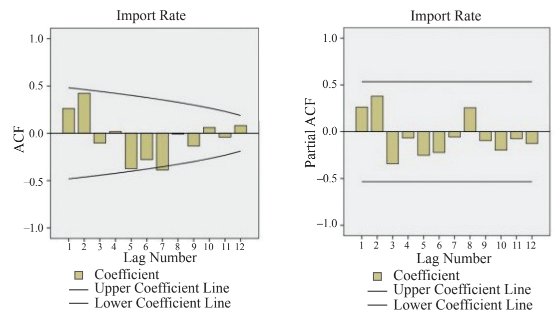

Figure 17: Graph of ACF and PACF of import rate.

$$ConsumptionR_t = 5.263 + (0.036 \times ConsumptionR_{t-1}) - (0.235 \times ConsumptionR_{t-2}) - (1.692 \times 10^{-7} \times Pig_t) + (7.613 \times 10^{-9} \times Chicken_t) \quad (8)$$

From Equation (8), the r-square value is 0.945 which means that this equation can explain the variation of the consumption rate at 94.5% with a mean absolute percent error at 2.042.

3.2.5 Import rate data with corn price and soybean price data

Consider the import rate time series data shown in Figure 16. This figure indicates that the import rate sharply decreased in 2001 and gradually increased from 2003 reaching a peak in 2008 and then fluctuating in a downward direction.

Considering the ACF and PACF for the corn production data in Figure 17, the model is indeterminate. Hence, the simulations to test for the best ARIMA (p,d,q) have been done and found that the appropriate time series model with the highest r-square value to forecast the import rate is the ARIMA (2,0,2) model. The results are demonstrated in Table 9 and can be developed equation as demonstrated in Equation (9).

From the analysis data in Table 9. It can be concluded that the Equation (11) can forecast the import rate by using previous values with an r-square of 0.307 and MAPE 232.256.

Table 9: Model fit statistics

Model Description						
					Model Type	
Model ID Import Rate Model_1					ARIMA(2,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.307	.307	.155	232.256	.	0	.
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
Import Rate Constant			.189	.071	2.654	.026
AR Lag 1			–.339	.456	–.745	.475
Lag 2			.060	.495	.122	.905
MA Lag 1			–.814	12.039	–.068	.948
Lag 2			–.993	29.480	–.034	.974

$$ImportRatet = 0.189 - 0.339\ ImportRatet_{-1} + 0.060\ ImportRate_{t-2} + 0.814\ a_{t-1} + 0.993\ a_{t-2} + a_t \quad (9)$$

To test the correlation function between the noise residual from import rate with corn price and soybean price. From Figure 18 it can be seen that there are no cross correlations among the noise residual from import rate with corn price and soybean price.

3.2.6 Corn price data with rain, plant acreage, temperature, consumption rate, number of pigs, number of chickens and import rate

This process is to test for correlation directly among noise residual of the corn price data with rain, plant acreage, temperature, consumption rate, number of pigs, number of chickens and import rate data. The results are indicated as in Figure 19.

From the cross-correlation function in Figure 19. It is clear that there are no factors correlated with the corn price data. The results of the analyses from 3.2.1 to 3.2.6 are shown in Table 10.

Table 10: Cross correlation analysis summarization

Group	Dependent Variable	Independent Variable	CCF
1	Corn Price	Corn Production	N
		Soybean Price	Y
2	Corn Production	Rain	N
		Plant Acreage	Y
		Temperature	N
3	Plant Acreage	Consumption Rate	N
		Import Rate	N
4	Consumption Rate	Amount of Pig	Y
		Amount of Chicken	Y
5	Import Rate	Corn Price	N
		Soybean Price	N
6	Corn Price	Rain	N
		Plant Acreage	N
		Temperature	N
		Consumption Rate	N
		Amount of Pigs	N
		Amount of Chickens	N
		Import Rate	N

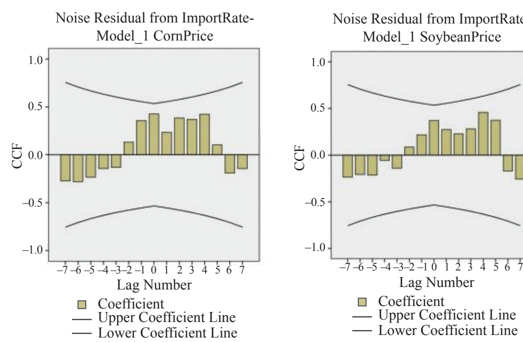

Figure 18: Graph of cross correlation function between noise residual from import rate with corn price and soybean price.

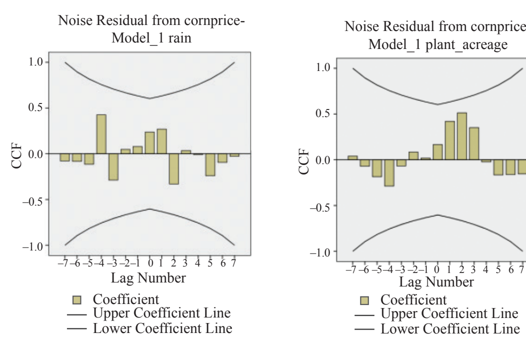

Figure 19: Graph of cross correlation function among the noise residual of corn price data with rain, plant acreage, temperature, consumption rate, number of pigs, number of chickens and import rate data.

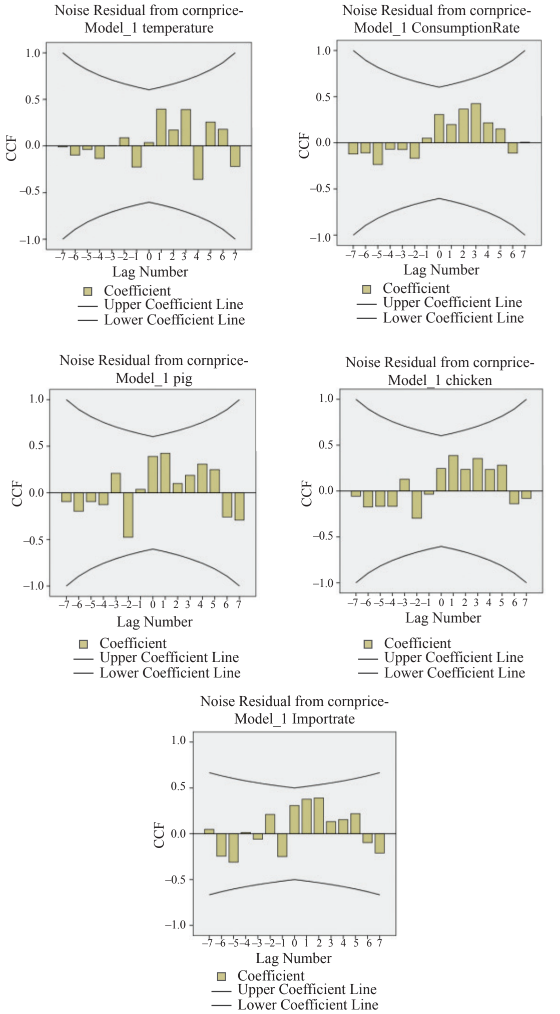

Figure 19: (*Continued*) Graph of cross correlation function among the noise residual of corn price data with rain, plant acreage, temperature, consumption rate, number of pigs, number of chickens and import rate data.

From Table 10, it can be concluded that corn price data have been affected only by the soybean price data solely. Hence, Equation (3) would be the appropriate forecasting equation for corn price data.

3.2.7 Soybean price time series data analysis.

From Equation (3), it is possible to know the soybean price in the prediction period (t). The soybean price data would then be analyzed. The soybean price time series data are plotted in Figure 20. This figure

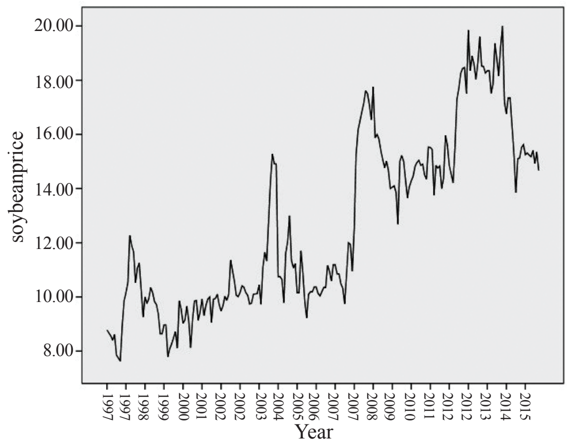

Figure 20: Graph of time series data of soybean price.

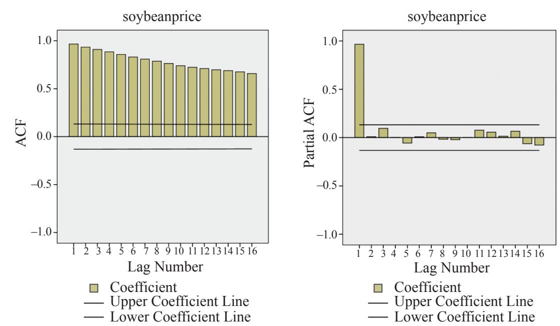

Figure 21: Graph of ACF and PACF of soybean price data.

demonstrates that the data comprise a trend variation. The autocorrelation function (ACF) and partial autocorrelation function (PACF) can be used to determine the correlation between the data and the time lag data itself. The ACF and PACF are illustrated in Figure 21.

As seen Figure 21, the ACF trends to decrease exponentially. The value of the PACF is significantly higher than zero at 1 lag significantly. Hence, the appropriate time series model to forecast the corn price is ARIMA (1,0,0) model. The results are demonstrated in Table 11.

From the forecasting model for ARIMA (p,d,q) in Equation (1). Hence, the ARIMA (1,0,0) forecasting model for soybean price from the model parameters is as shown in Equation (10).

$$Soybeanprice_t = 0.3944 + 0.969\ Soybeanprice_{t-1} \quad (10)$$

Hence, from Equations (3) and (10), the corn price forecasting model would be as given in Equation (11).

Table 11: Model fit statistics

Model Description						
					Model Type	
Model ID soybean price Model_1					ARIMA(1,0,0)(0,0,0)	
Model Statistics						
Model Fit Statistics					**Ljung-Box Q(18)**	
Stationary R-square	R-squared	RMSE	MAPE	Statistics	DF	Sig.
.935	.935	.836	4.648	10.456	17	.883
ARIMA Model Parameters						
			Estimate	**SE**	**T**	**Sig.**
soybean price Constant			12.507	1.541	8.115	.000
AR Lag 1			.969	.016	62.329	.000

$$Cornprice_t = 0.2098 + 1.161\ Cornprice_{t-1} - 0.196\ Cornprice_{t-2} + 0.068\ (0.3944 + 0.969\ Soybeanprice_{t-1}) \quad (11)$$

The next step is to forecast by using the second part of collected actual corn price data during January 2016 to December 2016. To test the forecasting performance, the forecasting data would be tested with Holt-Winters Multiplicative Seasonal (HWMS) model and Holt-Winters Additive Seasonal (HWAS) model. The forecasting values and fit graph from the ARIMA, the HWMS model and the HWAS model are demonstrated in Table 12 and Figure 22 respectively. The forecasting performance comparisons from 3 methods are shown in Table 13 which indicate that the ARIMA model generates the lowest forecasting errors of RMSE, MAPE and MSE during January 2016 to December 2016 of 0.8678, 12.1009 and 4.7592 respectively.

Table 12: Forecasting data results

Period	Actual Corn Price	Forecast		
		ARIMA	HWMS	HWAS
Jan-16	8.16	7.74	7.74	7.74
Feb-16	7.95	7.85	8	8
Mar-16	7.56	7.96	8.16	8.16
Apr-16	7.66	8.18	8.54	8.54
May-16	7.61	8.22	8.66	8.66
Jun-16	7.84	8.19	8.69	8.69
Jul-16	8.1	8.2	8.61	8.61
Aug-16	7.7	7.87	8.04	8.04
Sep-16	6.94	7.54	7.46	7.46
Oct-16	6.04	7.57	7.52	7.52
Nov-16	6.07	7.71	7.79	7.79
Dec-16	6.17	7.75	7.75	7.75

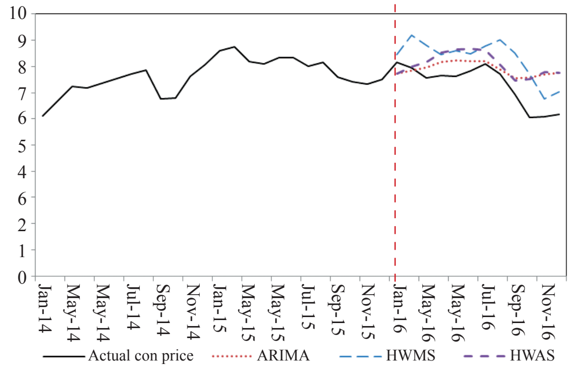

Figure 22: Graph of data fit values for forecasting data with 3 forecasting methods.

Table 13: Forecasting performance comparisons

Method	RMSE	MAPE	MSE
ARIMA	0.8678	12.1009	0.7531
HWMS	1.078	16.8836	1.1627
HWAS	0.9759	14.6492	0.9524

4 Conclusions

This objective of this research is to establish the corn price forecasting model, to forecast the corn price by using the Box-Jenkins model and compare forecasting performance with other methods. The results indicate that a corn price forecasting model by using only the two previous data points with current soybean price can generate an r-square of 0.951 with a mean absolute per cent error at 4.653. This means that the forecasting equation can explain the variation of corn price at 95.1% with mean absolute per cent error (MAPE) at 4.651.

The performance comparisons with the HWMS model and the HWAS model indicate that the ARIMA model can generate the lowest forecasting error. And the results of testing the ARIMA model with the actual corn price data during January 2016 to December 2016 generated RSME, MAPE and MSE values at 0.8678, 12.1009 and 0.7531 respectively.

Acknowledgments

The author would like to thank my advisor, Asst.Prof. Dr.Vichai Rungreunganun, for many helpful comments and for his encouragement to complete the research. My sincere gratitude to Field and Renewable Energy Crops Research Institute, Department of Agriculture, Dr.Chutima Koshawatana, Mr.Werapon Ponrukdee and team members for providing the author with the necessary information for this research.

References

[1] P. Milind and D. Isha, "Zea maize: A modern craze," *International Research Journal of Pharmacy*, vol. 4, no. 6, pp. 39–43, May 2013.

[2] D. Pimentel and T. W. Patzek, "Ethanol production using corn, switchgrass, and wood; biodiesel production using soybean and sunflower," *Natural Resources Research*, vol. 14, no. 1, pp. 65–76, Mar. 2005.

[3] K. Parker, M. Salas, and V. C. Nwosu, "High fructose corn syrup: Production, uses and public health concerns," *Biotechnology and Molecular Biology Review*, vol. 5, no. 5, pp. 71–78, Dec. 2010.

[4] E. Whittaker, (2014, Apr.). The Fundamental Factors Affecting Corn Price from 1982–2013. Tufts University, Massachusetts, USA [Online]. Available: https://ag.purdue.edu.agecorn

[5] P. C. Westcott and L. A. Hoffman, *Price Determination for Corn and Wheat: The Role of Market Factors and Government Programs*. Washington, DC: Technical Bulletin, 1999.

[6] N. Condon, H. Klemick, and A. Wolverton. (2013, Oct.). Impacts of Ethanol Policy on Corn Prices: A Review and Meta-analysis of Recent Evidence. National Center for Environmental Economics. Washington, DC, USA [Online]. Available: http://www.epa.gov/economics

[7] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.

[8] J. W. Forrester, "Industrial dynamics," *Science*, vol. 135, no. 3502, pp. 426–427, 1961.

[9] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*. New York: John Wiley & Sons, 1998.

[10] S. Akhtar and S. Rozi, "An autoregressive integrated moving average model for short-term prediction of hepatitis C virus seropositivity among male volunteer blood donors in Karachi, Pakistan," *World Journal of Gastroenterology*, vol. 15, no. 3, pp. 1607–1612, Apr. 2009.

[11] C. Christodoulos, C. Michalakelis, and D. Varoutas, "Forecasting with limited data: Combining ARIMA and diffusion models," *Technological Forecasting & Social Change*, vol. 77, pp. 558–565, Jan. 2010.

[12] M. O. Nasiru and S. O. Olanrewaju, "Forecasting airline fatalities in the world using a univariate time series model," *International Journal of Statistics and Applications*, vol. 5, no. 5, pp. 223–230, May 2015.

[13] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081.

[14] M. Williams and A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[15] W. Riansut and K. Thongrit, "Forecasting the prices of field corn in Thailand," *RMUTP Research Journal*, vol. 11, no. 1, pp. 1–14, Jan. 2017.

[16] M. Kerdsomboon and M. Varaphakdi, "Forecasting of agricultural products and prices," M.S. thesis, Department of Statistics, Faculty of Science, Chulalongkorn University, 1999 (in Thai).

[17] P. Kitworawut and V. Rungreunganun, "An application of analytical hierarchy process (AHP) for affect factor to corn price in Thailand market," *Journal of Advanced Agricultural Technologies*, vol. 4, no. 3, pp. 280–284, Sep. 2017.