**Forecasting the red lentils commodity market price using SARIMA models**

**Abstract**

Canada is the world’s largest producer of lentils, accounting for 32.8 % of total production in the world. However, the production of lentils will prone to fluctuate due to the impact of erratic factors such as weather condition and economic crises. Consequently, the price of the commodity will be changed and volatile. Therefore, the approach of modeling and forecasting future price based on the preceding data, will provide representative figures to make decisions regarding the lentils production for growers and end-users. Hence, the objective of this study is to model and forecast the red lentils prices using Seasonal Autoregressive Integrated Moving Average model (SARIMA). Eight years of weekly data starting from 2010 to 2019 which comprises 521 observations, obtained from Saskatchewan.ca were used in this study. The average red lentils price in Saskatchewan was dollar 24.75 per 100 pounds and weekly prices were highly fluctuating over the time. The seasonality and volatility of red lentils are modeled and forecasted by calculating the seasonal index and applying SARIMA models to the time series. The results reveal that the SARIMA(2,1,2)(0,1,1)[52] model provides the best in-sample and out-sample performance when predicting the red lentils prices. Hence, this model can be utilized by both growers and end-users in making optimal production decisions and in managing overall price risk.

**Keywords**: Forecasting, Lentils, Model, SARIMA, Seasonal Index.

1. **Introduction**

Humans consume protein in different procedures, including animal sources as meat and plant sources as pulses [1]. Particularly, lentils are a rich source of Protein, Carbohydrates, Fiber, etc. Further, one major benefit of pulse consumption over meat consumption is that pulses contain only tiny amounts of fat with compared to meat. Moreover, lentils also can be used for livestock feed. Table 1 shows the nutrition value of 100g of lentils. Lentils contain greater amount of Protein and Carbohydrate, whereas Fat content is low in lentils.

Table 1: Nutritional Value for every 100 g of lentils

|  |  |
| --- | --- |
| Nutrition | Content per 100 g of lentils |
| Energy(kcal) | 324 |
| Protein(g) | 25.4 |
| Fat(g) | 1.5 |
| Fiber(g) | 17.0 |
| Carbohydrate (g) | 44.8 |

 Source: FAO, 2019

Table 2 exhibits the essential micronutrients and vitamin B9 for each 100g of lentils. Lentils comprise higher amounts of Iron, Phosphorus, and Copper.

Table 2: Essential Micronutrients and B-9 Vitamin for every 100 g of lentils

|  |  |
| --- | --- |
| Nutrient | Content per 100 g of lentils |
| Iron(mg) | 7.1 |
| Magnesium(mg) | 66 |
| Phosphorus(mg) | 291 |
| Potassium(mg) | 752 |
| Zinc(mg) | 3.55 |
| Copper(mg) | 0.41 |
| B9 / Folate(mg) | 150 |

 Source: FAO, 2019

Canada is the major producer and exporter of lentils in the world as they exported approximately 2.03 million metric tons of lentils in the crop year 2018/19 to over hundred countries. They began the lentil production in 1970’s and currently there are over 5000 active lentil farmers in Canada. The province of Saskatchewan is the major contributor of 95% of the lentils production in Canada. Typically, lentils are planted in early May and harvested in mid-August in Saskatchewan.

However, the production of red lentils varies due to some influences such as weather, trade wars and financial policies etc. As a consequence, the red lentils price fluctuation is volatile which causes great impact to the growers, sellers, policy makers and consumers. There are complex relationships among influential factors. Thus, the precise forecasting is challenging. The current methods are mainly focused on qualitative analysis rather than the quantitative forecasting approach in literature. Some studies applied the artificial neural network and genetic algorithm without considering period of production.

The red lentils price is influenced by the season significantly. Hence, investigating the price fluctuation is needed the SARIMA model which accounts the seasonal effect in the time series. Subsequently, this study recommends a quantitative prediction method of red lentils price in Canada by applying SARIMA model in order to produce a decision-making tool for each associates.

* 1. **Literature Review**

Many of the studies have been conducted on forecasting future production and price of agricultural commodities specifically pulses based on the historical data by applying time series analysis.

Production of pulses in Kenya was forecasted using ARIMA model[2] indicated that ARIMA (1,1,2) model was the appropriate model to forecast the pulses production in Kenya and decreasing trend in the predicted production by 2030. Therefore, due to the increasing tendency in population growth, the estimated results produce a clue that there won’t be enough pulses to feed the growing population in Kenya by 2030.

One of the agricultural commodities, which is consumed by majority of the people can be identified as Cucumber. Forecasting vegetable price is also challenging task due to the seasonal variation. Hence, [3] used SARIMA model which considers the seasonal effect, to investigate an effective model of forecasting Cucumber price. SARIMA(1,0,1)(1,1,1)12 model was selected as the best fitted model which provides feasible short-term warning of vegetable price.

[4] focused on estimating price volatility of major pulses including lentils in India using GARCH model. High fluctuation of the production of pulses led to the high price variability in the market. Further, results emphasized that the volatility of price in the current period depends on the previous period.

1. **Methodology**
	1. **Materials and Methods**

The study is aimed to forecast price of red lentils in Saskatchewan, Canada. The study was carried out using weekly time series data for the period of 2010 to 2018. The data on red lentils was collected from askatchewan.ca, AGR Market Trends, Government of Saskatchewan, Canada. There were 521 observations.

The total number of 469 observations was used for model building and 52 observations among those 469 data points were used for assessing the in-sample performance. Remained 52 observations were used for performing out-sample performance.

* 1. **Model Identification**

In the identification stage the data is tested for stationary and the augmented dickey fuller (ADF) test and Phillips-Perron (PP) test were applied. Seasonal index were calculated to identify the seasonal pattern. Decomposed plots were used to identify the time series components; seasonal, trend, cyclic and random component in the data over time. Seasonality is represented by the seasonal component at time t. When a time series is influenced by seasonal factors there exists a seasonal pattern. Residual component describes the random or irregular influences at time t.

Non-Stationary time series data has statistical properties, which change with time. So, its required to change the data into stationary time series data by obtaining the first difference of the time series, before building the predictive model.

* + 1. **SARIMA**

Seasonal ARIMA model (SARIMA) is formed by adding seasonal terms in the ARIMA models.

$SARIMA\left(p,d,q\right)\left(P,D,Q\right)[S]$ (1)

Where p is a non-seasonal autoregressive order, P is a seasonal autoregressive order, q is a non-seasonal moving average order, Q is a seasonal autoregressive order, d and D are the order of common difference and seasonal difference.

SARIMA(p,d,q)(P,D,Q)[S] models are written as

$\left(1-ϕ\_{1}B^{ω}- ϕ\_{2}B^{2ω}- . . . - ϕ\_{p}B^{Pω}\right) × \left(1- φ\_{1}B- φ\_{2}B^{2}- . . . - φ\_{p}B^{p}\right) × \left(1- B^{ω}\right)^{D} \left(1-B\right)^{d} Q\_{n}\left(t\right)=\left(1- Θ\_{1}B^{ω}-Θ\_{2}B^{2ω}- . . . - Θ\_{1Q}B^{Qω}\right)×\left(1- θ\_{1}B- θ\_{2}B^{2}- . . . - θ\_{q}B^{q} \right)e(t)$ (2)

$ϕ$ is non-seasonal parameter of autoregression and θ is nonseasonal parameter of movingaverage, $φ$ is seasonal parameter of autoregression and Θ is seasonal parameter of moving average, ω is frequency and B is the differential variable.

The number of times the series is differenced determines the order of d. The AR and MA signatures are determined using non- seasonal and seasonal Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. A theoretical AR model of order p have an ACF that decays and a PACF that cuts off at lag p while a theoretical MA model of order q consist of a PACF that decays and an ACF that cuts off at lag q. The model with the minimum AIC and BIC values is selected as the model that fits the data best.

* 1. **Estimation of parameters and diagnostic checking**

Parameters of best fitted SARIMA model were estimated using Akaike information criterion(AIC) and Bayesian information criterion(BIC). Then the significance of the model parameters were assessed using t-test statistics. The residuals from the estimated model were generated and tested whether they resemble a white noise series (uncorrelated and have zero mean) by investigating ACF, PACF plot and performing Ljung-Box statistic test respectively. Heteroscedasticity of the residuals were detected using Autoregressive conditional heteroskedasticity Lagrange Multiplier (ARCH-LM) test.

If the parameter estimates were insignificant and the residual were not a white noise then the entire process of model identification, parameter estimation and diagnostic checking was repeated until the appropriate model was attained.

* 1. **Forecasting**

After selection of an appropriate model, future values of the time series were forecasted for in-sample and out-sample, and the confidence intervals for the forecasts were generated. Reliability of forecasted values based on selected model was checked by computing Sum of Squared Errors(SSE), Mean Absolute Error (MAE), Mean Squared Error(MSE), Root Mean Square Error (RMSE) and Theil’s inequality coefficient (TIC).

$SSE=\sum\_{t=1}^{n}\left(Y\_{t}- \hat{Y}\_{t}\right)^{2} $ (3)

$MAE=n^{-1}\sum\_{t=1}^{n}\left|σ\_{t}^{2}-\hat{σ}\_{t}^{2}\right| $ (4)

$MSE=n^{-1} \sum\_{t=1}^{n}\left(σ\_{t}^{2}-\hat{σ}\_{t}^{2} \right)^{2}$ (5)

$RMSE=\sqrt{\sum\_{t=1}^{n}\frac{\left(\hat{Y}\_{t}- Y\_{t}\right)^{2}}{n}}$ (6)

$TIC=\frac{\sqrt{\frac{1}{h+1}\sum\_{t=1}^{n}\left(Y\_{t}-\hat{Y}\_{t}\right)^{2}}}{\sqrt{\frac{1}{h+1} } \sum\_{t=1}^{n}\left(Y\_{t}\right)^{2} - \sqrt{\frac{1}{h+1} \sum\_{t=1}^{n}\hat{Y}^{2 }}}$ (7)

where  *n* is the number of forecasts, $σ\_{t}^{2}$and  $\hat{σ}\_{t}^{2}$ are the actual volatility and the volatility forecasts obtained from SARIMA  models respectively.

1. **Results and Discussion**

According to the summary statistics in table 3, the maximum price of red lentils in Saskatchewan was dollar 52.28 per 100 pounds in 2nd week, 2016 while the lowest red lentils price was dollar 14.25 per 100 pounds in 33rd week, 2018. On average red lentils price in Saskatchewan from 2010 to 2018 was dollar 24.75 per 100 pounds. Though, a significant increase in red lentils price was realized from 47th week of 2015 to 23rd week of 2016, afterwards the price was declined as illustrated in figure 1. Further, the time series plot indicated that the lentils price is highly fluctuating over the time.

Table 3: Descriptive statistics of the lentils price during 2010 and 2018

|  |  |
| --- | --- |
| Minimum | 14.25 |
| Maximum | 52.28 |
| Median | 21.78 |
| Mean | 24.06 |
| Q1 | 18.16 |
| Q3 | 27.5 |
| Mode | 33.75 |



Figure 1: Red Lentils Price in dollars per 100 pounds in Saskatchewan (2010-2019)

The seasonal indices for weekly red lentils price were calculated (Appendix A) and plotted. Figure 2 exhibits maximum price in the 22nd Week and minimum price in the 37th Week of the year. Further, seasonal indices were negative from 10th week to 14th week and 28th week to 48th week whereas within first 10 weeks, from 15th week to 27th week and last four weeks in the year seasonal indices were positive. It emphasized that the lentils price exhibits seasonality, since the price is low during the harvesting period while the price increases through planting period.

Figure 2: Weekly Seasonal Index Plot for red lentils price from 2010 - 2019

ACF and PACF plots for the original red lentils price data were shown in Figure 3. The results implied that price data was not stationary since the ACF die off slowly. Further, ADF test and PP test for red lentils price data was performed to confirm whether the data was stationary or not. The p-values of ADF and PP tests which were 0.696 and 0.7212 respectively, were greater than 0.05. Therefore time series was not stationary at 5% significance level. Hence, the data was differenced to make it stationary. The p-values of ADF and PP test were 0.01 which was less than 0.05. It indicates that the differenced data was stationary at 5% level of significance. The first differencing was sufficient to make the data stationary, hence price was integrated of order one (d=1). Figure 4 shows a plot of the differenced red lentils price data against time.



Figure 3: ACF and PACF plots for red lentils price



Figure 4: Red Lentils Differenced data

In order to obtain the order of non-seasonal and seasonal AR and MA, the Plots of ACF and PACF for the non-seasonal and seasonal differenced price data were obtained and results were shown in Figure 5 and Figure 6. There were insignificant spikes in all plots. After analyzing those plots, based on the seasonal and non-seasonal AR and MA orders, six parsimonious models were selected for the model building purpose.



Figure 5: ACF and PACF plots for Non-Seasonal Differenced Data

Figure 6: ACF and PACF plots for Seasonal Differenced Data

The table 4 indicates the selected best models according to AIC and BIC values. Accordingly, SARIMA (1,1,1)(0,1,1)[52] has the minimum AIC and BIC value.

Table 4: AIC and BIC values of selected models

|  |  |  |
| --- | --- | --- |
| Model | **AIC** | **BIC** |
| SARIMA(1,1,1)(0,1,1)[52] | 2.94851 | 2.98303 |
| SARIMA(1,1,1)(0,1,2)[52] | 2.95270 | 2.99586 |
| SARIMA(1,1,2)(0,1,1)[52] | 2.95278 | 2.99593 |
| SARIMA(2,1,1)(0,1,1)[52] | 2.95278 | 2.99594 |
| SARIMA(2,1,2)(0,1,1)[52] | 2.95320 | 3.00498 |
| SARIMA(3,1,3)(0,1,2)[52] | 2.96419 | 3.04187 |

However, according to the in-sample and out-sample performance, SARIMA(2,1,2)(0,1,1)[52] which had the lowest SSE, MAE, MSE, RMSE and TIC, was selected as the feasible model. Further ACF and PACF plots of residuals and Ljung-Box test statistics indicated that the residuals of the selected model were random, white noise and independent. Then, ARCH-LM test was performed to assess the heteroscedasticity of residuals. Since the p-value was 0.999 which was greater than 0.05, residuals SARIMA(2,1,2)(0,1,1)[52] model was not heteroscedastic at 5% significance level.

Table 5: Accuracy measurements of SARIMA(2,1,2)(0,1,1)[52]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | Model | SSE | MAE | MSE | RMSE | TIC |
| Out-sample | SARIMA(2,1,2)(0,1,1)[52] | 146.583 | 1.386 | 2.820 | 1.679 | 0.002 |
| In-Sample | 66.275 | 0.913 | 1.275 | 1.129 | 0.003 |

According to the parameter estimation results in table 6, SARIMA(2,1,2)(0,1,1)[52] model can be expressed as following equation (9).

Table 6: Parameter estimation of SARIMA(2,1,2)(0,1,1)[52] model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   | Estimate | SE | t\_value | p\_value |
| ar1 | -0.0594 | 0.2873 | -0.2067 | 0.8364 |
| ar2 | 0.4882 | 0.1745 | 2.7981 | 0.0054 |
| ma1 | 0.2103 | 0.2849 | 0.738 | 0.4609 |
| ma2 | -0.4423 | 0.1625 | -2.7226 | 0.0068 |
| sma1 | -0.9998 | 0.201 | -4.9752 | 0 |

$\left(1-ϕ\_{1}B\right) × \left(1- φ\_{2}B^{2}\right) × \left(1- B^{52}\right)(1-B) Q\_{n}\left(t\right)=\left(1- θ\_{2}B^{2}\right)e(t)$ (8)

$\left(1+0.9998\*B\right) × \left(1+0.4423\*B^{2}\right) × \left(1- B^{52}\right)(1-B) Q\_{n}\left(t\right)=\left(1- 0.4882\*B^{2}\right)e\left(t\right)$ (9)



Figure 7: The forecasted red lentils price using SARIMA(2,1,2)(0,1,1)[52] in Canada

Lentils prices from January to December 2019 were predicted using best fitted SARIMA(2,1,2)(0,1,1)[52]. The forecasted prices within 80% and 95% prediction intervals are shown in figure 7. Forecasted values in 2019 (Appendix B) were shown fluctuating pattern and decreasing trend with respect to the price in last week of December, 2018.

1. **Conclusion**

Red lentils are one of the major pulses which comprise high nutritional value. Therefore, consumer’s demand is high on lentils in many countries. However, due to the impact of many factors, price of lentils is fluctuating. Hence, farmers, policy makers, and traders are interested in forecasting lentils prices to attain optimum marketing decisions and to cope with price risk. In this study, Seasonal ARIMA modeling was used to forecast the price of red lentils in Saskatchewan, Canada who is the major contributor in the lentils export market. The best fitted model for price was identified as SARIMA(2,1,2)(0,1,1)[52]. Consequently, this model can be applied as a short-term decision making tool on lentils price. Since the price is volatile, for long-term forecasting the model should be modified by adding new actual values and regular monitoring of price should be done by the relevant authorities.

**Acknowledgement**

Authors are grateful to Mr. Chandika Jayawardane, Chief Analyst & Managing Director, Global Trade Finance and Quantitative Research | WISEWEL LLC, Kandy, Sri Lanka, for providing the data.

**References**

[1]Lakkakula P., Olson F. and Ripplinger D.,” Pea and Lentil Market Analysis”, Final Research Report. Agribusiness and Applied economics. North Dakota state University. July 2017.

[2]Esther N.M., & Magdaline N.W., “ARIMA Modeling to Forecast Pulses Production in Kenya”, Asian Journal of Economics, Business and Accounting. Vol.2, No. 3, February 2017, pp. 1-8.

[3]Luo C, Zhou L & Wei Q., “Application of SARIMA Modeling in Cucumber Price Forecast”, Applied Mechanics and Materials. Trans Tech Publications, Switzerland. Vol. 373-375, August 2013, pp. 1686-1690.

[4]Bisht A. & Kumar A., “Estimating Volatility in Prices of Pulses in India: An application of GARCH Model”, Economic Affairs. Vol. 64, No. 3, September 2019, pp. 513-516.

[5]Alibhutto M.C. & Ariyarathna H.R., “Forecasting Weekly Temperature Using ARIMA Model”, A Case Study for Trincomalee In Sri Lanka. International Journal of Science, Environment and Technology, Vol. 8, No 1, 2019, 124 – 131.

 [6]Chen P., Niu A., Liu D., Jiang W. & Ma B., “Time Series Forecasting of Temperatures Using SARIMA”, An Example from Nanjing. IOP Conference Series: Materials Science and Engineering. Vol. 394, August 2018, (052024).

[7]Kaur H & Ahuja S. (2019). “ARIMA Modeling for Forecasting the Electricity Consumption of a Health Care Building”, International Journal of Innovative Technology and Exploring Engineering (IJITEE). Vol. 8, No. 12, pp. 2795-2799.

[8]Rawal V. & Navarro, D. K., “The Global Economy of Pulses”, Food and Agriculture Organization of the United Nations. Rome.2019. (<http://www.fao.org/3/i7108en/I7108EN.pdf>)

**Appendix A: Weekly Seasonal Index**

|  |  |
| --- | --- |
| Week Number | Index |
| 1 | 0.292545 |
| 2 | 0.427481 |
| 3 | 0.908048 |
| 4 | 0.923774 |
| 5 | 0.986862 |
| 6 | 0.575174 |
| 7 | 0.983507 |
| 8 | 0.891285 |
| 9 | 0.472652 |
| 10 | -0.43159 |
| 11 | -0.54031 |
| 12 | -0.5216 |
| 13 | -0.6451 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Week Number | Index |  | Week Number | Index |  | Week Number | Index |
| 14 | -0.06802 |  | 27 | 0.354961 |  | 40 | -1.06758 |
| 15 | 0.323817 |  | 28 | -0.46093 |  | 41 | -1.08688 |
| 16 | 0.771755 |  | 29 | -0.68058 |  | 42 | -1.10501 |
| 17 | 1.246274 |  | 30 | -0.79214 |  | 43 | -1.16193 |
| 18 | 1.53637 |  | 31 | -0.54089 |  | 44 | -1.16755 |
| 19 | 1.321904 |  | 32 | -0.83907 |  | 45 | -1.17383 |
| 20 | 1.548486 |  | 33 | -0.901 |  | 46 | -1.39826 |
| 21 | 1.779233 |  | 34 | -0.98062 |  | 47 | -1.50631 |
| 22 | 1.956926 |  | 35 | -1.12846 |  | 48 | -0.18161 |
| 23 | 1.732225 |  | 36 | -1.09895 |  | 49 | -0.03438 |
| 24 | 1.305772 |  | 37 | -1.60569 |  | 50 | 0.041263 |
| 25 | 1.028389 |  | 38 | -0.96866 |  | 51 | 0.496114 |
| 26 | 0.978892 |  | 39 | -1.12038 |  | 52 | 0.323635 |

**Appendix B: Out sample Performance**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Actual | Forecast |  | Date | Actual | Forecast |  | Date | Actual | Forecast |
| 1/2/2019 | 18.09 | 18.3236 |  | 5/8/2019 | 17.38 | 18.0548 |  | 9/11/2019 | 15.53 | 15.3466 |
| 1/9/2019 | 18.34 | 19.0623 |  | 5/15/2019 | 18.16 | 18.2318 |  | 9/18/2019 | 15.53 | 15.1702 |
| 1/16/2019 | 18.84 | 18.8828 |  | 5/22/2019 | 18.16 | 18.2663 |  | 9/25/2019 | 16.06 | 15.1972 |
| 1/23/2019 | 18.91 | 18.862 |  | 5/29/2019 | 18.19 | 17.851 |  | 10/2/2019 | 16.21 | 15.1542 |
| 1/30/2019 | 19.16 | 18.1343 |  | 6/5/2019 | 18.5 | 17.4289 |  | 10/9/2019 | 16.5 | 15.1157 |
| 2/6/2019 | 19.53 | 17.8787 |  | 6/12/2019 | 18.5 | 17.2713 |  | 10/16/2019 | 17.5 | 15.0394 |
| 2/13/2019 | 18.97 | 17.8681 |  | 6/19/2019 | 18.5 | 17.3515 |  | 10/23/2019 | 17.86 | 15.0177 |
| 2/20/2019 | 18.97 | 17.528 |  | 6/26/2019 | 18.5 | 16.9384 |  | 10/30/2019 | 17.86 | 14.9982 |
| 2/27/2019 | 18.97 | 16.4708 |  | 7/3/2019 | 18.5 | 16.1219 |  | 11/6/2019 | 18 | 14.7602 |
| 3/6/2019 | 18.97 | 16.4047 |  | 7/10/2019 | 18.31 | 15.8632 |  | 11/13/2019 | 18.36 | 14.6388 |
| 3/13/2019 | 17.53 | 16.6121 |  | 7/17/2019 | 18.31 | 15.7123 |  | 11/20/2019 | 18.36 | 15.9525 |
| 3/20/2019 | 17.22 | 16.6068 |  | 7/24/2019 | 17.66 | 15.927 |  | 11/27/2019 | 18.36 | 16.0916 |
| 3/27/2019 | 17.22 | 17.1377 |  | 7/31/2019 | 17.38 | 15.5961 |  | 12/4/2019 | 18.5 | 16.1623 |
| 4/3/2019 | 17.03 | 17.313 |  | 8/7/2019 | 17.25 | 15.5108 |  | 12/11/2019 | 18.57 | 16.6106 |
| 4/10/2019 | 18.34 | 17.5929 |  | 8/14/2019 | 17.03 | 15.4133 |  | 12/18/2019 | 18.86 | 16.4321 |
| 4/17/2019 | 17.38 | 18.0878 |  | 8/21/2019 | 16.28 | 15.2457 |  | 12/25/2019 | 18.86 | 16.3936 |
| 4/24/2019 | 17.38 | 17.9934 |  | 8/28/2019 | 15.97 | 15.256 |  |  |  |  |
| 5/1/2019 | 17.38 | 17.8303 |  | 9/4/2019 | 15.97 | 14.7307 |  |  |  |  |