Forecasting Trends in Corn Prices in Illinois, Iowa, and Nebraska

Abstract

Corn is vital to the health of the world economy and is the lifeblood of the American Midwest. The ability to predict corn prices plays a major role in corn market trends. Previous research has yielded effective models for corn price prediction, but no previous work has considered the sensitivity of corn prices to time lags or geography. This project attempts to fill that gap by using regression analysis on inflation-adjusted corn price data to create a multiple linear regression-based prediction model using available data from Midwestern states during the timeframe when corn is grown. The prediction model displays an R^2 value of 0.7642 and a one-year time lag was used for prediction due to the unavailability of data. Future research should focus on gathering data for corn prices and predictor variables in different time resolutions.

Background and Introduction

Corn is the most widely-produced grain in the world, and given its integral role in the agricultural, food, gasoline, chemicals, and plastics industries, it is no wonder why corn gets all the attention it does. America is especially interested in the health of its corn production given the Midwest's place as the world's largest corn breadbasket and its outsized influence on the Midwestern economy. Without corn, the Midwest would vanish in an instant and bring America's agricultural sector and economy down with it.

To ensure that corn stays lucrative and maximizes total agricultural profits, farmers need to know how much corn to plant and sell in any given year. That requires the ability to anticipate future corn prices, a subject covered by a copious amount of literature. Many papers on corn price modeling have proposed models that fit existing corn data very well (R^2 coefficients > 0.9), but little research has been done on corn price forecasting for periods longer than 1 month, accounting for the role of inflation in prediction, or considering how variables at the state level affect corn prices (Xu, 2020; Xu, 2018; Goodwin & Schnepf, 2000; Kitworawut & Rungreunganun, 2019; Henrique et al., 2019; Ge & Wu, 2020). This paper attempts to fill those gaps in research by using regression analysis on inflation-adjusted corn price data to create a multiple linear regression model using available data from the state and national levels.

Data and Exploratory Analysis

Since the focus was on forecasting future corn prices, the study chose to analyze the states with the most corn production: Iowa, Illinois, and Nebraska.[1] Due to the significant price fluctuation during the growth period, the study also chose to analyze three specific months: May, July, and September, which are the corn planting season, mid-season, and harvesting season, respectively. The data collection process consisted of looking through government-sourced or publicly available information through well-known institutions, which included USDA (United States Department of Agriculture), FRED (Federal Reserve Economic Data), etc. A common issue that was run into was the variation of units due to mismatched scales, with certain data only being available in monthly format and others only being available in yearly format. However, it is possible to work around it by grouping variables based on their respective units. Since the dataset contains data on a wide year range (1980-2009), all variables with units in US dollars were adjusted for inflation by indexing prices to 1980 dollars using the CPI. From this point onwards, the paper refers to corn prices, crude oil prices, and subsidies by their inflation-adjusted values. There were 19 potential variables to predict corn prices: month, state, year, acreage, corn production, precipitation, interest rate, crude oil price, subsidies, corn consumption, change in subsidies, change in acreage, change in production, change in consumption, change in precipitation, change in interest rate, change in crude oil price, previous corn price (time = t-1), and change in previous corn price. "Change in" variables were calculated using $x_{i,t} = (\frac{x_{i,t}}{x_{i,t-1}} - 1) \cdot 100$, where $x_{i,t}$ denotes the predictor variable in its regular scale at a year t.

Since the goal of this paper is to predict future corn prices, data from year $t-1$ is used to estimate corn prices in year t. Due to the time lag necessary to "initialize" the data, data from 1980 and 1981 are omitted. We created visualizations to study the relationship between the 19 predictor variables and corn price.

The side-by-side boxplot of corn price versus month shows a slight right-skewed distribution among all months, with May having a higher median value along with right-skewed outliers for May and September (Figure A). The boxplot between state and corn price reveals a slight right-skewed distribution, with Iowa having a higher median than Illinois or Nebraska of approximately $1.30 and no clear outliers (Figure A). The histogram of corn acreage shows an unimodal graph with a slight left-skewed and the scatterplot shows a cluster of data points

[1] Please refer to the Appendix for additional information on the figures used in this paper.

between 10000-13000 corn acreage (Figure B). The distribution of corn production is roughly symmetrical, and there is a very weak negative association between corn production and price (Figure C). The distribution of precipitation data is right-skewed, and there is no association between precipitation and corn price (Figure D). In regards to interest rates, there is an unimodal bar graph with most interest rates being between 0-10% (Figure E). Crude oil follows an unimodal, slightly left-skewed distribution, and the scatterplot between corn price and crude oil shows a cluster of data at $5-10 (Figure F). Regarding corn subsidies, there is an unimodal distribution with slight outliers on the upper ends and a slight negative association with corn prices (Figure G). Lastly, corn consumption has an unimodal right-skewed distribution and a slight negative association with corn prices (Figure H). A comprehensive collection of EDA plots is contained in the Appendix.

Table 1. Variable codebook. Outlines predictor and response variables.

Variable Name	Variable Description	Valid Range or Variable Code
Month	Month of given data	May, July, September
State	State from where data is from	Illinois, Iowa, Nebraska
Year (time=t)	Year from which data is from	1982-2009
Acreage (time=t-1)	Yearly Acres of Corn Harvested (thousands) in State	~5000-1500
CornProduction (time=t-1)	Bushels of corn harvested (thousands) in State	~700000-2500000
Precipitation (time=t-1)	Monthly precipitation in state (inches)	~0-10
InterestRate (time=t-1)	United States interest rate %	~0-17
CrudeOilPrice (time=t-1)	Inflation adjusted crude oil price (USD)	~5-40
Subsidies (time=t-1)	Inflation adjusted US yearly aggregate agriculture subsidies (Billion USD)	~1-13
Consumption (time=t-1)	Yearly aggregate corn consumption in the US (Million bushels)	~700-5100
CornPrice (time=t)	Inflation adjusted corn price (USD/Bushel)	0-4

Model and Results

A multiple linear regression model was used to explain the variance in corn prices as the corn prices have continuous values and moderate linear relationships were observed between some of the predictors and the response variable. To improve the fit of the model and resolve violations of linear regression model assumptions, the response variable (corn prices) was transformed to the negative square root as suggested by the Box-Cox function (Figure J). After the transformation was applied, the diagnostic plots used to verify the multiple linear regression assumptions had improved. Initially, violations of the linearity, constant variance, and normality assumptions of a linear regression model were found. Specifically, the QQ-Plot of the regular data revealed deviations from linearity which improved after the transformation (Figure L). The "Residual versus Predicted" plot before the transformation revealed a widening spread of residuals as $\hat{y}$ increased, which violates the constant variance assumption. Constant variance improved after applying the transformation (Figure K). Several predictor variables exhibited non-linear relationships with corn prices. This improved after applying the transformation to corn prices. The transformation was kept in the final model as it greatly improved the fit of the multiple linear regression model.

Since the original dataset contained a large number of variables, the study used a subset of the potential predictors for the final model due to the complexity associated with using the majority of the predictor variables. The best subset of variables for the model was found by performing forward, backward, and both-direction stepwise selection. Backward and both-direction selection suggested the same model. In contrast, forward selection suggested a more complex model with more predictor variables that could be reduced to the model suggested by backward and both-direction selection through a nested F-test. Therefore, the model suggested by backward and both-direction selection was chosen. An additional predictor variable (Yearly % change in subsidies) was safely removed through a nested F-test due to a high p-value associated with the predictor in the model. The final model contained 12 predictor variables. The resulting least squares regression line (LSRL) for corn prices is

$$\widehat{CornPrice_t}^{-1/2} = -46.27 + (2.338*10^{-2})Year_t + (3.211*10^{-5})Acreage_{t-1} + (9.636*10^{-8})CornProduction_{t-1} + (-8.625*10^{-3})InterestRate_{t-1} + (-3.305*10^{-3})CrudeOilPrice_{t-1} + (2.328*10^{-2})Subsidies_{t-1} + (-1.204*10^{-4})Consumption_{t-1} + (-1.199*10^{-3})AcreageChange_{t-1} + (0.1347)Price_{t-1} + (-1.854*10^{-3})PriceChange_{t-1} + (-0.0179)I_{May} + (0.0489)I_{Sept} + (-0.096)I_{Iowa} + (0.132)I_{Nebraska}$$

where

$$I_{May} = \begin{cases} 1 & Month = May \\ 0 & Month = July \end{cases} \quad I_{Sept} = \begin{cases} 1 & Month = September \\ 0 & Month = July \end{cases}$$

$$I_{Iowa} = \begin{cases} 1 & State = Iowa \\ 0 & State = Illinois \end{cases} \quad I_{Nebraska} = \begin{cases} 1 & State = Nebraska \\ 0 & State = Illinois \end{cases}$$

Table 2. Summary of Model Coefficients for $\widehat{CornPrice}^{-\frac{1}{2}}$

Variable	Estimate Slope	P-Value	95% Confidence Interval
Intercept	-4.62E+01	<2E-16	(-5.45E+1, -3.79E+1)
Year	2.34E-02	<2E-16	(1.92E-2, 2.75E-2)
Acreage	3.21E-05	0.0015	(1.23E-5, 5.18E-5)
CornProduction	9.64E-08	0.0018	(3.61E-8, 1.52E-7)
InterestRate	-8.62E-03	0.00653	(-1.48E-2, -2.43E-3)
CrudeOilPrice	-3.30E-03	0.0362	(-6.39E-3, -2.128E-4)
Subsidies	2.33E-02	<2E-16	(1.84E-2, 2.81E-2)
Consumption	-1.20E-04	4.16E-10	(-1.56E-4, -8.40E-5)
AcreageChange	-1.20E-03	0.013	(-2.14E-3, -2.54E-4)
Price (t-1)	1.35E-01	7.70E-08	(8.68E-2, 1.82E-1)
PriceChange (t-1)	-1.85E-03	6.24E-11	(-2.38E-3, -1.32E-3)
I_May	-1.79E-02	0.103	(-3.95E-2, 3.69E-3)
I_Sept	4.89E-02	2.45E-05	(2.64E-2, 7.12E-2)
I_Iowa	-9.60E-02	3.31E-09	(-1.26E-1, -6.52E-2)
I_Nebraska	1.31E-01	8.90E-06	(7.44E-2, 1.88E-1)

The final model successfully predicts future corn prices because it has an $R^2_{Adj} = 0.7642$. It explains 76.42% of the variance in corn prices, the highest of the models found by forward, backward, and both-direction stepwise selection. The value of $\hat{\beta}_{Subsidies}$ means that every dollar increase in subsidies is associated with a 0.0233 increase in $\widehat{CornPrice}^{-\frac{1}{2}}$, and the p–value suggests a significant relationship. We are 95% confident that a marginal monthly increase in $\hat{\beta}_{Subsidies}$ is associated with a increase in $\widehat{CornPrice}^{-\frac{1}{2}}$ between 0.0184 and 0.0281, *ceteris paribus.*

Discussion/Conclusions

This study attempted to construct a prediction model for monthly inflation-adjusted corn prices at the state level using a combination of predictors at the state, national, monthly, and yearly levels. The high R^2 value suggests that the model consistently predicts corn prices at close to their actual value for any given month. However, time-series data is notoriously difficult to fit a multiple linear regression on due to the issues it causes for statistical interpretation. The study was unable to find the optimal amount of time lag in the time-series data due to a lack of available data for many of the predictor variables. Due to this constraint, the time lag was arbitrarily chosen to be one year, but future research could involve finding the optimal time lag for predicting future corn prices. The model may be a better fit with an optimal time lag as more variance could be potentially explained.

Of all the predictors used in the final model, price was most sensitive to year, federal subsidies, corn consumption, and previous changes in corn prices. There are plausible reasons why each of these predictors is useful for corn price forecasting. The year that corn price data was collected and the change in the price of corn from the previous year possibly highlight external trends in the economy not accounted for by the other predictor variables used. Federal subsidies are a major source of revenue for farmers and can greatly influence the supply side of the corn market since subsidies play into the decision-making process that farmers make every year about their planting decisions. Therefore, the government can directly affect the price level of the corn market by changing how much they subsidize farmers. Likewise, the demand side of the corn market is determined by corn consumption, and if corn consumption directly affects the corn market, then it also directly affects the price of corn, which is consistent with the findings.

Given the way that the study set up the conditions behind the model, it is only useful for predicting corn prices in Illinois, Iowa, and Nebraska. Future research could entail expanding the analysis to a wider range of states or building new models to analyze international corn prices.

References

Federal Reserve Economic Data Economic Research (2023). *Government Subsidies: Federal: Agricultural.* https://fred.stlouisfed.org/series/L312041A027NBEA

Ge, Y., Wu, H. Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Comput & Applic* **32**, 16843–16855 (2020). https://doi.org/10.1007/s00521-018-03970-4

Goodwin, B. K. and Schnepf, R. (2000). *Determinants of Endogenous Price Risk in Corn and Wheat Futures Markets.* Journal of Futures Markets, 20(8), 753–774. doi:10.1002/1096-9934(200009)20:8<753::aid-fut3>3.0.co;2-f

Henrique, M., Ribeiro, D. M., Henrique, V., Ribeiro, A., Reynoso-Meza, G., dos, L., & Coelho, S. (2019). *Multi-Objective Ensemble Model for Short-Term Price Forecasting in Corn Price Time Series. 2019 International Joint Conference on Neural Networks (IJCNN).* doi:10.1109/ijcnn.2019.8851880

Integrated Public Use Microdata Series. *National Historical Geographic Information System.* https://www.nhgis.org/

Iowa State University. *Iowa Cash Corn and Soybean Price (USDA NASS).* https://www.extension.iastate.edu/agdm/crops/pdf/a2-11.pdf

Kitworawut, P. and Rungreunganun, V. *Corn Price modeling and forecasting using Box-Jenkins model.* Applied Science and Engineering Progress, vol. 12, no. 4, pp. 264–276, Oct.–Dec. 2019. https://ph02.tci-thaijo.org/index.php/ijast/article/view/232590/159014

United States Department of Agriculture Economic Research Service (2023). *Feed Grains: Yearbook Tables.* https://www.ers.usda.gov/data-products/feed-grains-database/feed-grains-yearbook-tables.aspx

United States Department of Agriculture National Agricultural Statistics Service. *USDA/NASS QuickStats.* https://quickstats.nass.usda.gov/

University of Nebraska-Lincoln Institute of Agriculture and Natural Resources. Historical Prices [Excel file]. https://cropwatch.unl.edu/economics/corn. Retrieved from Accessed December 11, 2023.

Xu, X. (2018). *Using Local Information to Improve Short-Run Corn Price Forecasts*. Journal of Agricultural & Food Industrial Organization, 16(1), 20170018. https://doi.org/10.1515/jafio-2017-0018

Xu, X. (2020). *Corn Cash Price Forecasting.* Amer. J. Agr. Econ., 102: 1297-1320. https://doi.org/10.1002/ajae.12041

Appendix

Figure A. EDA Testing for Categorical Variables (Month, State)

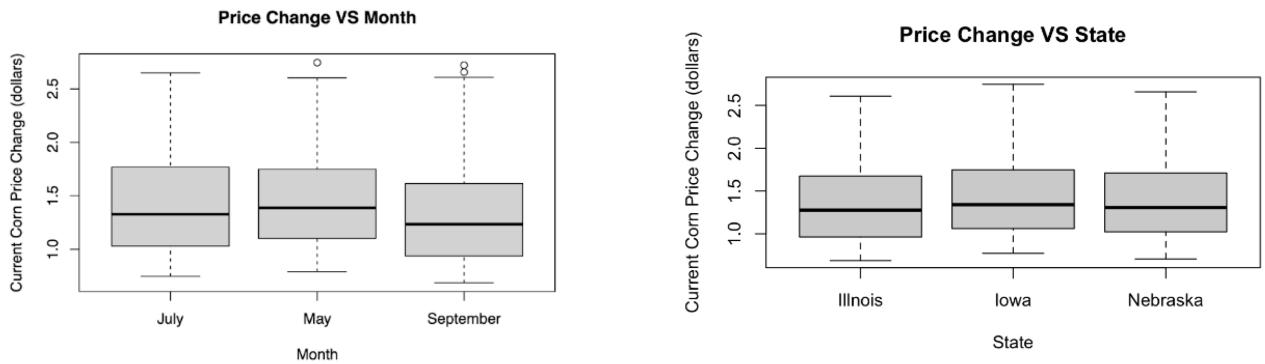

Figure B. EDA Testing for Corn Price vs Corn Acreage

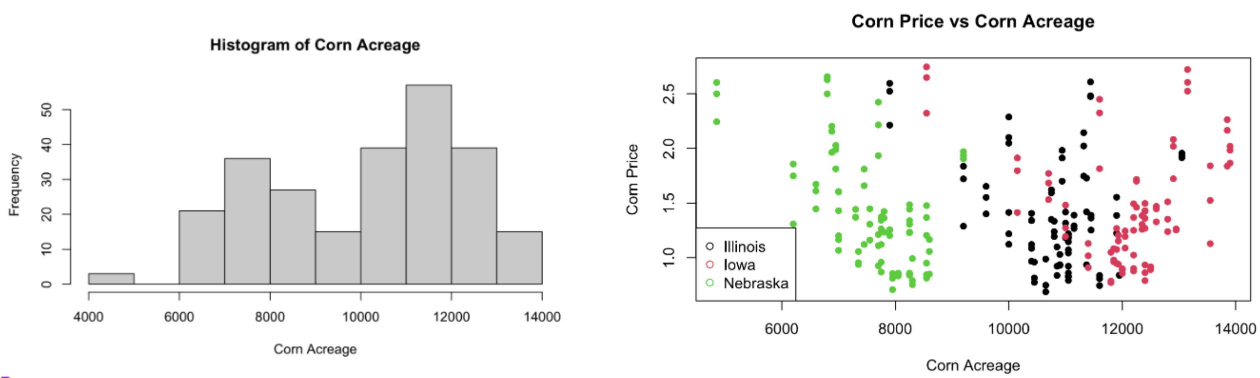

Figure C. EDA Testing for Corn Price vs Corn Production

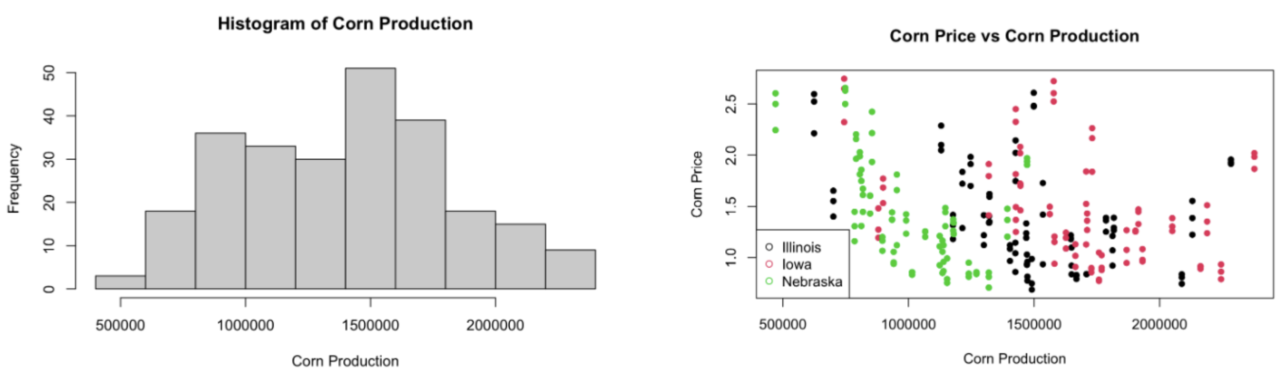

Figure D. EDA Testing for Corn Price vs Precipitation

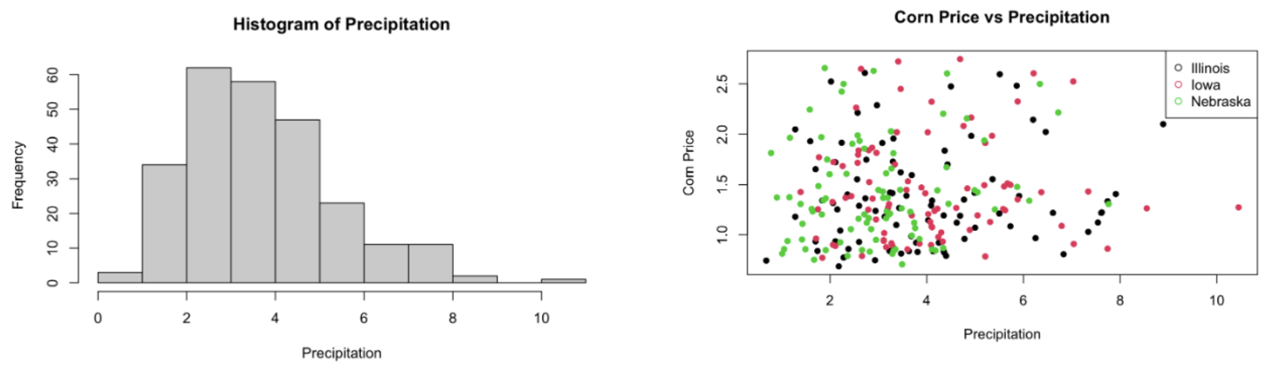

Figure E. EDA Testing for Corn Price vs Interest Rate

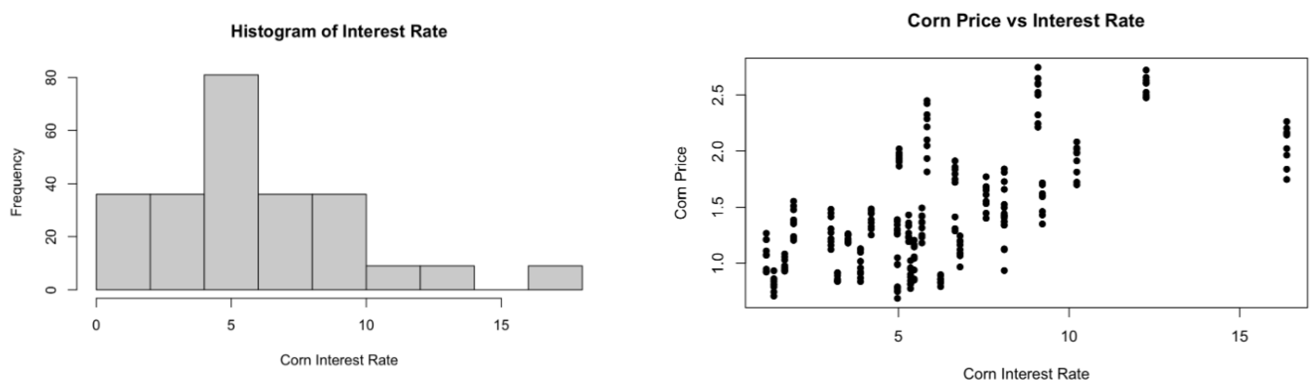

Figure F. EDA Testing for Corn Price vs Crude Oil Price

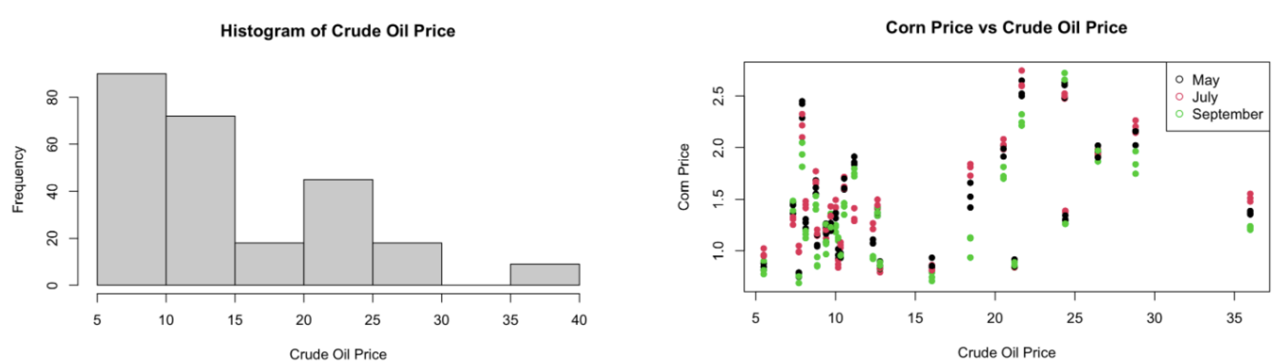

Figure G. EDA Testing for Corn Price vs Corn Subsidies

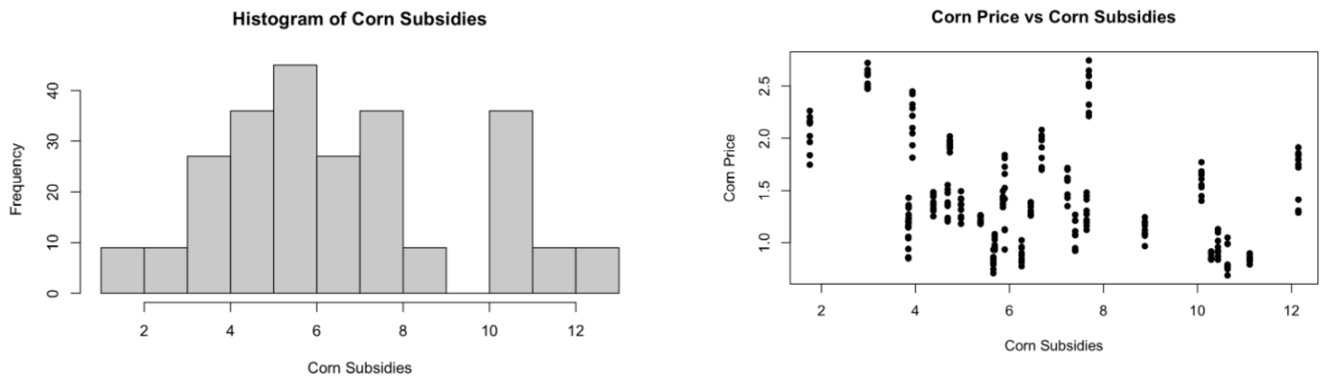

Figure H. EDA Testing for Corn Price vs Corn Consumption

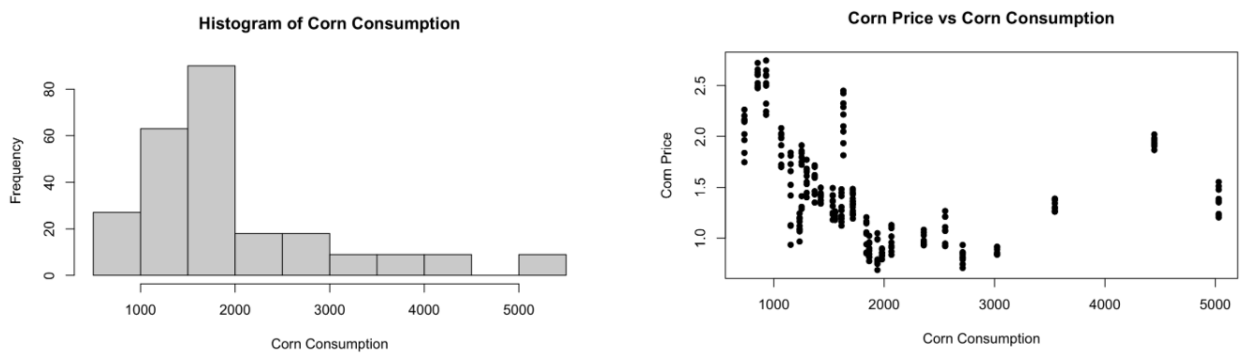

Figure I. Pairs Graph with the Predictor Variables

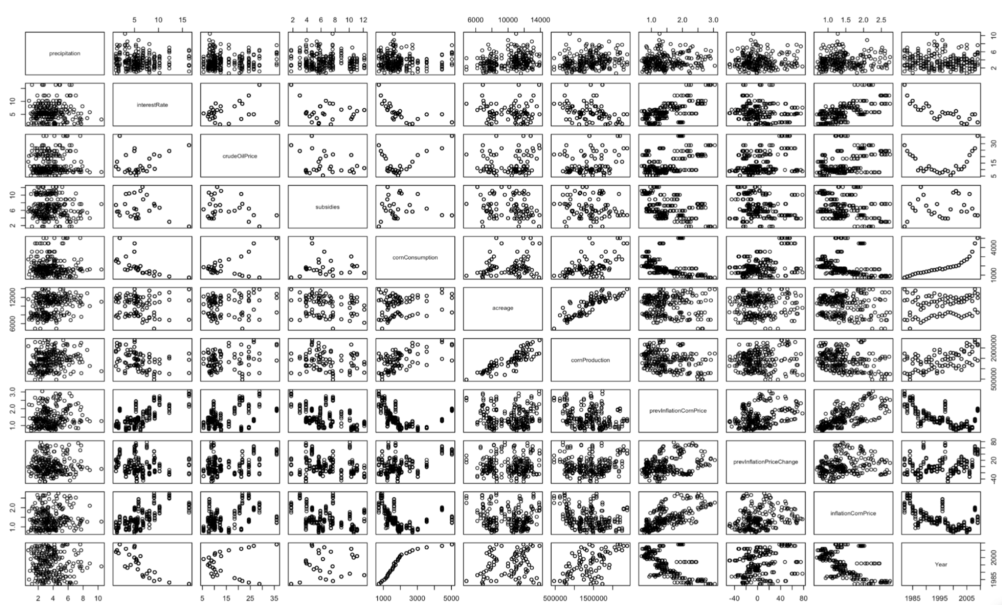

Figure J. Boxcox function for Corn Price

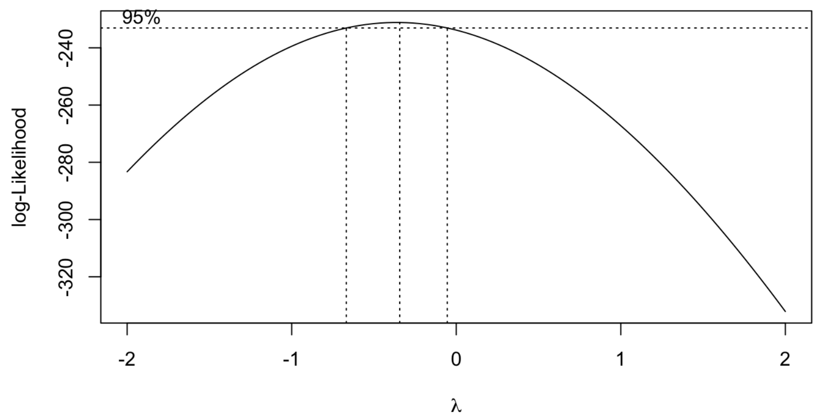

Figure K.

Diagnostic plots for Multiple Linear Regression Model for $\widehat{CornPrice}_t^{-1/2}$.