Price Optimization of Rice in Sri Lanka

for the Bachelor of Science Honours Degree in
Financial Mathematics and Industrial Statistics

By

C.S.N.Fernando
SC/2020/11822

Supervisor:
Dr. D.M. Samarathunga

Department of Mathematics
University of Ruhuna
Matara.

2024

DECLARATION

I, Name, declare that the presented project report titled," Price Optimization of Rice in Sri Lanka"is uniquely prepared by me based on the group project carried out under the supervision of Dr. D.M. Samarathunga Department of Mathematics, Faculty of Science, University of Ruhuna, as a partial fulfillment of the requirements of the level III Case Study II course unit MFM3151 of the Bachelor of Science Honours Degree in Financial Mathematics and Industrial Statistics in Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka. It has not been submitted to any other institution or study program by me for any other purpose.

Signature:.................................
Date:..................

SUPERVISOR'S RECOMMENDATION

I/We certify that this study was carried out by Name under my/our supervision.

Signature:.................................
Date:..................
Supervisor,
Department of Mathematics,
Faculty of Science,
University of Ruhuna.

Acknowledgment

In preparation for our case study, I received invaluable help and guidance from several respected persons, to whom I deserve my deepest gratitude. As I take pleasure in completing this report, first I would like to show my sincere gratitude to Dr. D.M. Samarathunga, from the Department of Mathematics at the Faculty of Science, University of Ruhuna for providing me with good guidance and support for the case study through numerous consultations. Furthermore, I extend my appreciation to all those who have directly and indirectly guided me in writing this project report.

Additionally, I would be thankful to the Head of the Department and the entire staff members of the Department of Mathematics for their assistance during the Case Study I Course Unit. Their support has been necessary in my progress. Many people, especially my group members, have given valuable comments and suggestions on my report which inspired me to improve the quality of the project report.

Contents

List of Figures

Chapter 1

Introduction

In Sri Lanka, Rice product is one of the main products and staple foods. Rice is the staple food of 21.8 million Sri Lankans and is the livelihood of more than 1.8 million cultivators. More than 30 percent of the total labor force is directly or laterally involved in the rice sector. Rice is cultivated in all sections of Sri Lanka during two thunderstorm seasons. They're called the Maha season and the Yala season. Maha Season starts in September and ends in March during the Northeast thunderstorm, and Yala season starts in May and ends in August. Price of the rice in Sri Lanka varied grounded on factors like variety and region. Typically, the average price for a kilogram of rice ranges from LKR 80 to LKR 150. Still, prices might have changed since also due to various economic, seasonal, and market influences.

1.1 Background of the study

The study focuses on optimizing the price of rice in Sri Lanka, a key staple food in the country. Sri Lanka has a significant dependence on rice both as an optimistic staple and as a profitable commodity. Similarly, the pricing of rice directly impacts the livelihoods of consumers, farmers, and various stakeholders within the agricultural sector.

1.1.1 Background Context

Rice is the primary food crop in Sri Lanka and plays a key role in the country's food security. It is a staple in the diet of the majority of the population and contributes significantly to the agricultural sector's output.

Sri Lanka's agricultural landscape is diverse, with various regions specializing in rice cultivation. Such as climate changes, soil types, irrigation systems, and farming practices influence rice production in the country.

The pricing of rice is influenced by a multitude of factors, including production costs, market demand, supply chain inefficiencies, international trade, weather conditions, government policies, and consumer preferences.

Fluctuations in rice prices can have significant socio-economic implications. For instance, higher rice prices might affect consumers' purchasing power, while low prices could impact the income of rice farmers, potentially affecting their livelihoods.

1.2 Aims

In our case study, The main aim of the study is to develop a suitable model for optimizing the price of rice in Sri Lanka. This involves understanding the complex factors influencing pricing dynamics and devising strategies to optimize price-setting mechanisms.

1.2.1 Objectives

1. To support the agricultural sector's sustainability in Sri Lanka. This includes offering equitable compensation to rice farmers, boosting agricultural output, and creating an environment that is more stable and profitable for individuals who cultivate rice.
2. To improve the rice industry's market efficiency. To do this, supply chains must be streamlined, inefficiencies must be eliminated, and a more responsive and transparent market structure must be established that benefits both producers and consumers.
3. To support and strengthen food security initiatives. Stable and optimized rice prices can ensure consistent access to affordable and nutritious food for the population.

4. To give stakeholders and policymakers important information. Offering suggestions on possible modifications to laws, farming methods, pricing schemes, or market restrictions that could have a favorable effect on the rice business is one way to do this.
5. To anticipate market changes, make informed decisions, and implement proactive measures to mitigate potential risks or capitalize on opportunities.

1.3 Literature Review

The foundation for understanding the economic drivers of rice prices is set out by Pingali and Rosegrant's (1995) [2] classification of rice as an inferior product in Asia. This perspective is accompanied by Mohanty's 2005 [5] analysis of agricultural price policies in India and provides insights that may be relevant to Sri Lanka's rice sector. and marketing and processing constraints (Weerahewa and Ward, Research specific to Sri Lanka delves into rice market stabilization policies (Jayasuriya and Shand, 2016) 2006). These studies provide insight into issues in the country's rice market and lay down possible strategies to improve them.

The theoretical frameworks for price optimization, as presented by Nguyen [6]and Deo 2018 [1] and Ray and Cheong 2015, provide useful information on how to formulate an efficient rice pricing strategy in Sri Lanka.

To understand how pricing strategies can respond to the diverse preferences of Sri Lanka consumers, a key factor is consumer behavior, and Caswell's (2017) research on customer demand for quality and Carpio MassandIsengildina's (2009) study about Consumer Preferences are helpful in this regard. The foundation for the development of a strong pricing model that takes into account different constraints and objectives can be found in theoretical knowledge on mathematics optimization techniques, Hillier and Lieberman, 2013, Taha (2016) [3]. The potential of machine learning in the rice industry is explored by Vargas-Guzman and García-Hernández (2020) in predicting rice yield, and Mishra and Misra's (2019) [4] review on the application of machine learning in agriculture suggests the value of integrating these algorithms into pricing models for enhanced predictive capabilities. A comprehensive understanding of market developments and dynamics in Sri Lanka's rice sector is possible through the introduction of Big Data Analytics, which was reviewed by Rajkumar and Nithya (2017) and Rehman et al. (2018).

Chapter 2

Material and Methods

2.1 Research Questions

1. What factors influence the optimal pricing of rice in the Sri Lankan market?
2. how can a balanced pricing strategy be determined to ensure both economic sustainability for rice producers and affordability for consumers?

2.2 Methodology

The purpose of the research is to create an equation to calculate a rice price depending on several variables. After that find an optimum price for the rice using some operational research approaches.

- The conceptual model:

  Here is the general conceptual model representing broader categories of predictors under the case study.

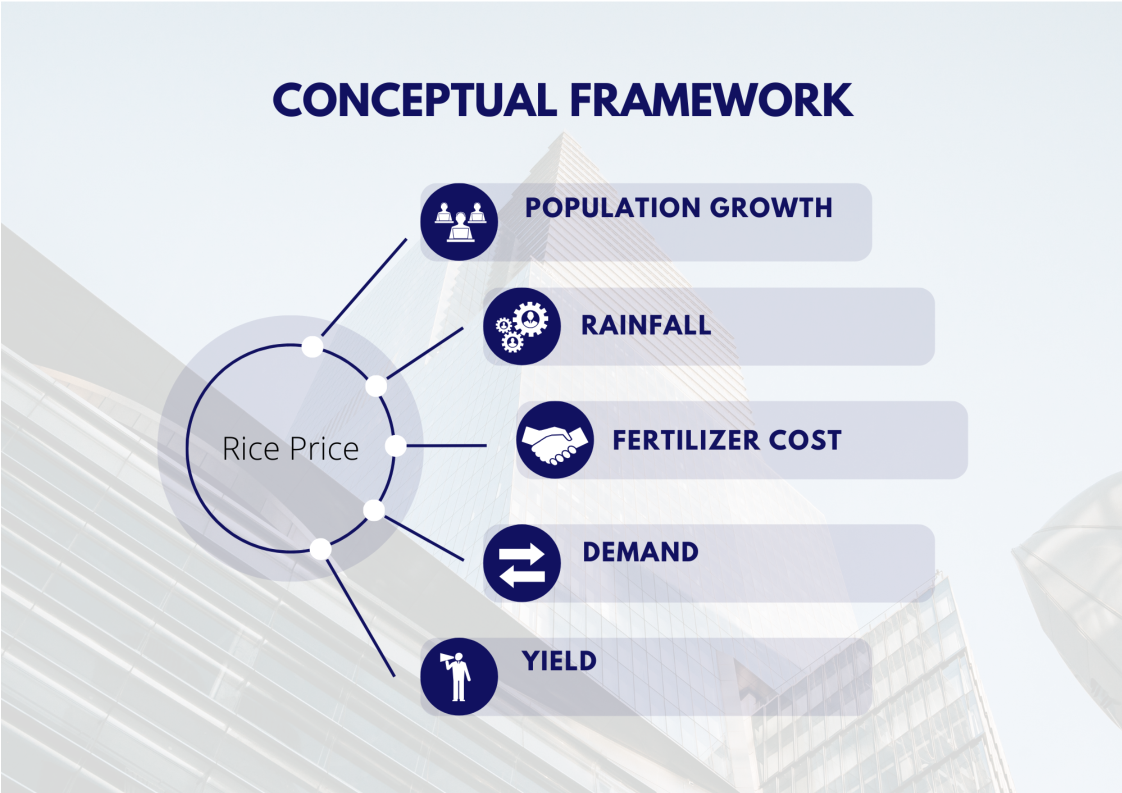

Figure 2.1: General Conceptual Model

- Research Design:

This is a quantitative case study that adopts multiple linear regression. There are multiple independent variables, yield, fertilizer cost, rainfall, population growth, and demand. Multiple linear regression allows researchers to examine the simultaneous effects of these variables on the dependent variable (Rice price) while controlling for potential confounding factors.

2.2.1 Ordinary Least Square Method

We can draw any number of lines through the data points. The set of data such that $(x_1, y_1), (x_2, y_2), \ldots.., (x_n, y_n)$. But we have to find the best fit for the data. To find the best-fitted line the most commonly used method is least square method.

Model : $y_i = a + bx_i + \epsilon_i$; i=1,2,....n, ϵ_i = residual

For given value of x_i there will be a difference between the value y_i and the corresponding value as determined by the line. We denote the difference by ϵ_i which is referred to as a deviation error or residual and may be positive, negative or zero.

According to the least square method ;

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{2.1}$$

2.3 Multiple Linear Regression Method

The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables. Regression models are used to describe relationships between variables by fitting a line to the observed data.

Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + +\beta_k x_{ki} + \epsilon_i;$ i=1,2,..n, ϵ_i = residual

According to the above research approach method we're using, our studies are connected around the 'Multiple linear regression' method. Also, this method is known as one of the most powerful methods in statistics for prediction. As an example, let's take our dependent variable 'Rice price' as 'y'. Let's take our independent variables as 'Yield(x1)', 'Fertilizer cost (x2)', 'Rainfall (x3)', 'Population growth (x4)', and 'Demand (x5). So, we can express our variables according to the multiple linear regression method. i.e

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon_i;$

Chapter 3

Data

3.1 Data Cleaning Process

- Dataset:

  A data set file needed for our case study was generated from R software. The data set has 101 observations with 7 variables. (An extract of the data set and instructions to load the uploaded data set to R studio is in the appendix)

- Data dictionary:

```
str (data)  # Check the data types of each column

## 'data. Frame':    101 obs. of  7 variables:
##  $ date              : char  "01/10/2023" "02/10/2023" "03/10/2023" "04/10/2023"
...
##  $ rice_price        : num  81.7 88.1 80.1 87.1 91.2 ...
##  $ yields            : num  3818 3577 4596 3212 3451 ...
##  $ fertilizer_cost   : num  50 50 50 50 50 ...
##  $ rainfall          : num  77.4 70.1 71.8 66.5 71.1 ...
##  $ population_growth : num  1.62 1.08 1.77 1.15 1.09 ...
##  $ premium_rice_demand: num  -8.16 28.74 25.56 6 29.17 ...
```

```
Summary (data) # Get summary statistics for each variable

##      date              rice_price         yields     fertilizer_cost
##  Length:100         Min.   : 74.77   Min.   :2560   Min.   :50.01
##  Class :character   1st Qu.: 87.98   1st Qu.:3560   1st Qu.:50.26
##  Mode  :character   Median : 93.84   Median :3941   Median :50.51
##                     Mean   : 94.17   Mean   :4024   Mean   :50.51
##                     3rd Qu.: 98.64   3rd Qu.:4491   3rd Qu.:50.75
##                     Max.   :120.30   Max.   :6334   Max.   :51.00
```

```
##     rainfall       population_growth premium_rice_demand
##  Min.   : 62.54   Min.   :-0.0644   Min.   :-8.155
##  1st Qu.: 76.52   1st Qu.: 1.0814   1st Qu.:15.328
##  Median : 85.50   Median : 1.4271   Median :22.126
##  Mean   : 86.11   Mean   : 1.3943   Mean   :20.881
##  3rd Qu.: 95.00   3rd Qu.: 1.6985   3rd Qu.:28.525
##  Max.   :113.47   Max.   : 2.5727   Max.   :45.970
```

Figure 3.1: Data Dictionary: Variable Type, Summary

Missing Data:

The summary and the context of missing data have been investigated. In our case, we tried three methods for handling the missing data. They are,

1. Dropping the observations with missing values.
2. Fill in the missing value from the median of the column for a given country.
3. Multiple Imputation.

date	0
rice_price	0
yields	0
fertilizer_cost	0
rainfall	0
population_growth	0
premium_rice_demand	0

Figure 3.2: No. of Missing Values

From the above figure, we can see that all variables are free of missing values.

- Preparation for analysis:

  The dataset includes 7 variables. All variables are numerical. Before doing the analysis, the variables should be filtered and mutated considering their observed properties in initial data preparation.

1. Checking Outliers.

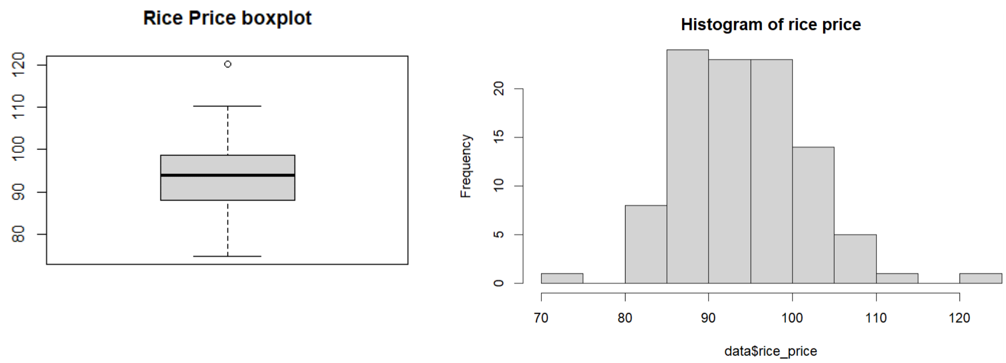

Figure 3.3: Visualization of Outliers

The above figure shows that one outlier has been observed in the dependent variable. The distribution of the dependent variable seems to be relatively symmetric. At this stage, the outliers have not been removed since the visualized outliers in the boxplot cannot be considered misinformation as we removed all noticed unrealistic observations previously. Several linear models will be constructed during analysis with and without outliers; eventually, the best model for our objectives will be selected.

Descriptive statistics have been used to check the validity of data.After removing, 100 observations of 7 numerical variables have been remained in the data set.

Chapter 4

Results

4.1 Exploratory data Analysis

Exploratory Data Analysis provides the foundations for rigorous and meaningful research; hence it is a very important phase. It helps investigators develop a deeper understanding of their data, gain early insights, and select the most relevant statistical models and methodologies to use for the study that follows.

1. **Summary Statistics:**

   We have got the numerical summary of the dataset to get an initial overview. This summary statistics includes mean, median, standard deviation, minimum, first quartile, median, third quartile, and maximum values for our variables.

```
     date             rice_price         yields      fertilizer_cost    rainfall       population_growth
 Length:100         Min.   : 74.48   Min.   :1356   Min.   :50.01   Min.   : 59.09   Min.   :0.660    
 Class :character   1st Qu.: 87.83   1st Qu.:3541   1st Qu.:50.26   1st Qu.: 76.27   1st Qu.:1.181    
 Mode  :character   Median : 94.99   Median :3993   Median :50.51   Median : 84.87   Median :1.493    
                    Mean   : 94.21   Mean   :3988   Mean   :50.51   Mean   : 86.85   Mean   :1.539    
                    3rd Qu.: 99.96   3rd Qu.:4471   3rd Qu.:50.75   3rd Qu.: 97.03   3rd Qu.:1.793    
                    Max.   :111.30   Max.   :5594   Max.   :51.00   Max.   :114.26   Max.   :2.843    
 premium_rice_demand
 Min.   :-0.0365    
 1st Qu.:14.2197    
 Median :21.0633    
 Mean   :20.6211    
 3rd Qu.:26.2318    
 Max.   :46.4461    
```

Figure 4.1: Summary statistic

2. **Visualization of Distribution of variables using Box-plots:**

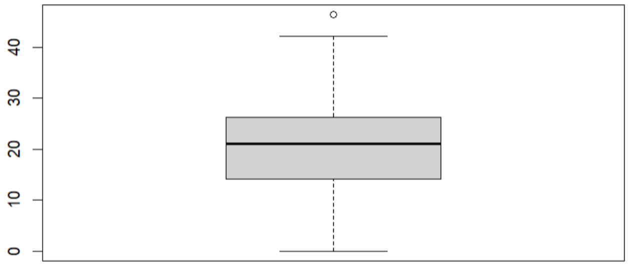

Figure 4.2: The box-plot of demand variable

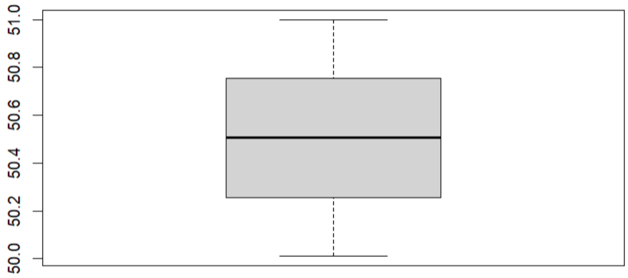

Figure 4.3: The box-plot of fertilizer cost variable

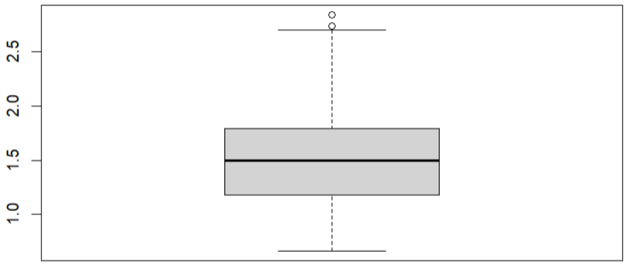

Figure 4.4: The box-plot of population growth variable

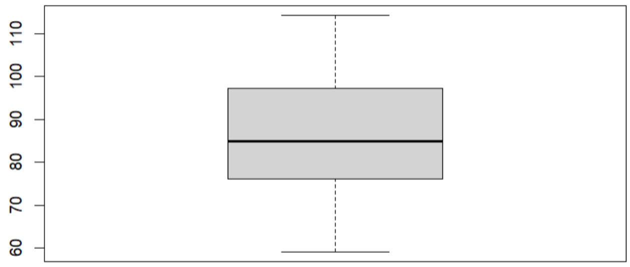

Figure 4.5: The box-plot of rainfall variable

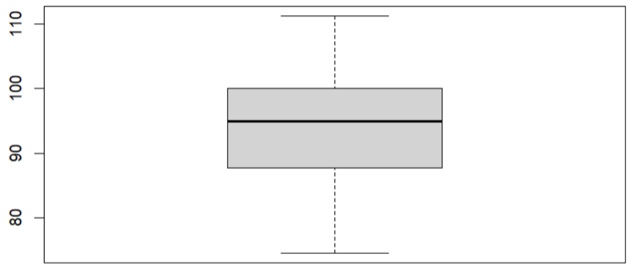

Figure 4.6: The box-plot of rice price variable

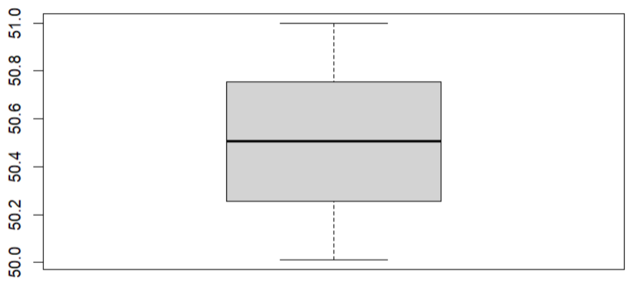

Figure 4.7: The box-plot of yields variable

According to the above figures, we can conclude that all variables are symmetrically distributed and there are some outliers.

3. **Scatter plots:**

   We can find the relationship between two variables using a scatter plot. Here are scatter plots to find a relationship between our dependent variable and independent variables.

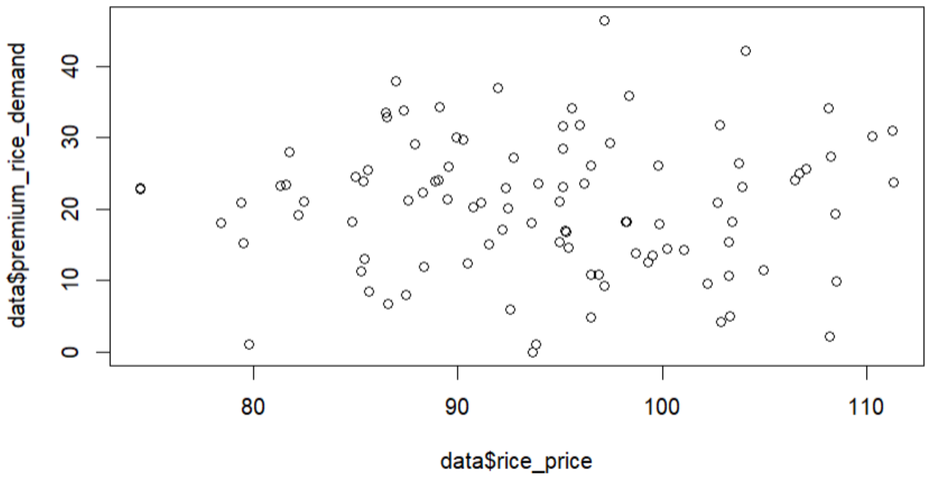

Figure 4.8: Demand and rice price

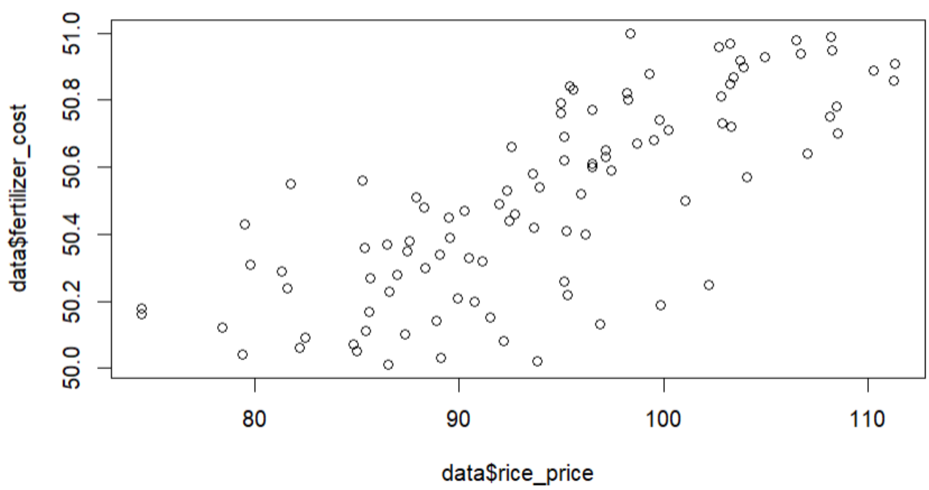

Figure 4.9: Fertilizer cost and rice price

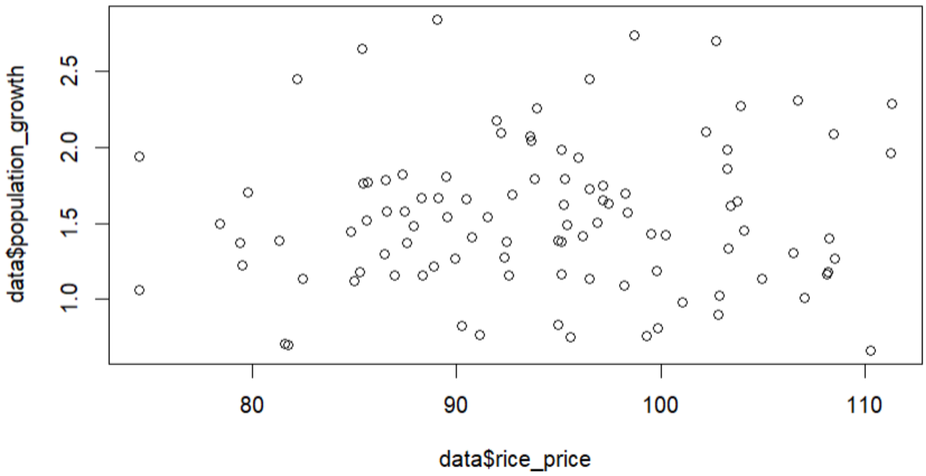

Figure 4.10: Population growth and rice price

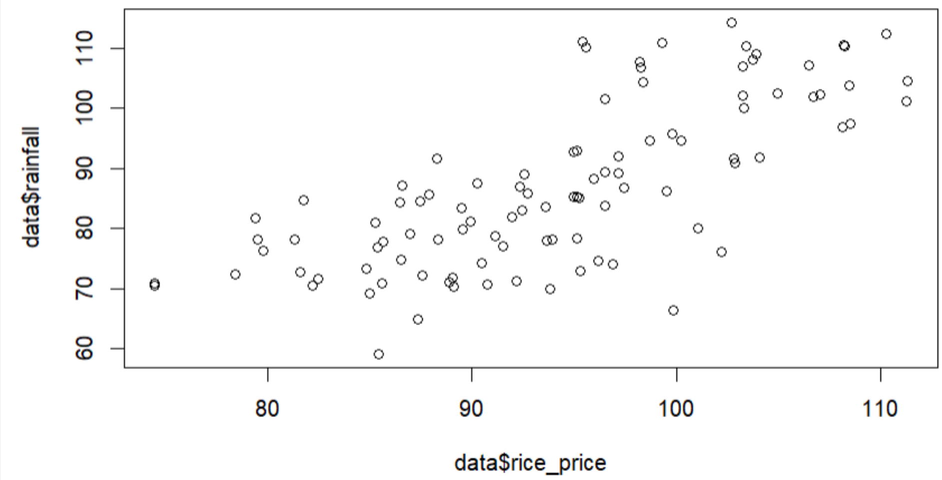

Figure 4.11: Rainfall and rice price

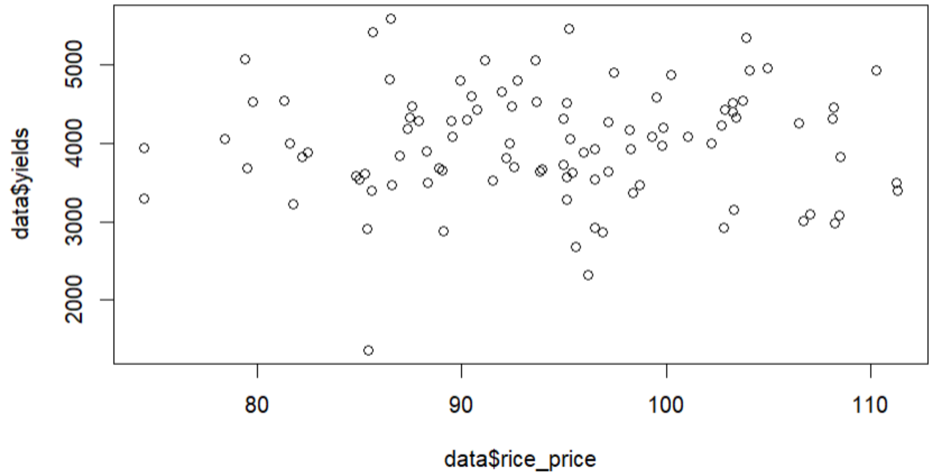

Figure 4.12: yield and rice price

We can conclude that the relationship between rice price vs fertilizer cost and also rainfall variables has a positive relationship, but we cannot see any relationship between other variables.

4.2 Quantitative analysis

Here we decide to create the initial multiple linear regression model by using all the available 4 numerical independent variables to predict rice price using the Ordinary Least Squares (OLS) method.

4.2.1 Interpretation of estimated coefficients:

```
Call:
lm(formula = rice_price ~ yields + rainfall + fertilizer_cost +
    population_growth, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-15.3930  -3.2807   0.6043   3.0781   9.0749

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -9.243e+02  2.024e+02  -4.566 1.50e-05 ***
yields            -6.113e-02  6.306e-02  -0.969    0.335
rainfall          -7.171e-03  9.659e-02  -0.074    0.941
fertilizer_cost    2.021e+01  4.148e+00   4.872 4.47e-06 ***
population_growth  6.089e-01  1.051e+00   0.579    0.564
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.537 on 94 degrees of freedom
Multiple R-squared:  0.6301,    Adjusted R-squared:  0.6144
F-statistic: 40.04 on 4 and 94 DF,  p-value: < 2.2e-16

Multiple R-squared: 0.630138
Adjusted R-squared: 0.6143992
```

Figure 4.13: Initial Regression Model

1. **Intercept**: The intercept, -924.3, is the predicted rice price when all other variables are equal to zero. This is not a meaningful value in this context, as it is not possible to have zero yields, rainfall, fertilizer cost, or population growth.

2. **Yields**: The coefficient for yields is -0.06113. This means that, on average, for each one-unit increase in yields, the predicted rice price decreases by 0.06113 units, holding all other variables constant. This suggests that there is a negative relationship between yields and rice prices. This could be due to several factors, such as increased supply leading to lower prices, or lower quality rice from higher yields fetching a lower price.

3. **Rainfall**: The coefficient for rainfall is -0.00717. This means that, on average, for each one-unit increase in rainfall, the predicted rice price decreases by 0.00717 units, holding all other variables constant. This suggests that there is a negative relationship between rainfall and rice prices. This could be due to several factors, such as increased rainfall leading to flooding and crop damage, or lower quality rice from wet conditions fetching a lower price.

4. **Fertilizer cost**: The coefficient for fertilizer cost is 0.2021. This means that, on average, for each one-unit increase in fertilizer cost, the predicted rice price increases by 0.2021 units, holding all other variables constant. This suggests that there is a positive relationship between fertilizer cost and rice price. This is likely because higher fertilizer costs lead to higher production costs, which are passed on to consumers in the form of higher rice prices.

5. **Population growth**: The coefficient for population growth is 0.6089. This means that, on average, for each one-unit increase in population growth, the predicted rice price increases by 0.6089 units, holding all other variables constant. This suggests that there is a positive relationship between population growth and rice prices. This is likely because increased demand from a growing population leads to higher rice prices.

4.2.2 Interpreting Adjusted R-Squared:

The adjusted R-squared of 0.6144 indicates that approximately 61.44% of the variation in rice price can be explained by the linear regression model. This is a moderately high value, suggesting that the model is an approximately good fit for the data.

4.2.3 Check for significance and model fit:

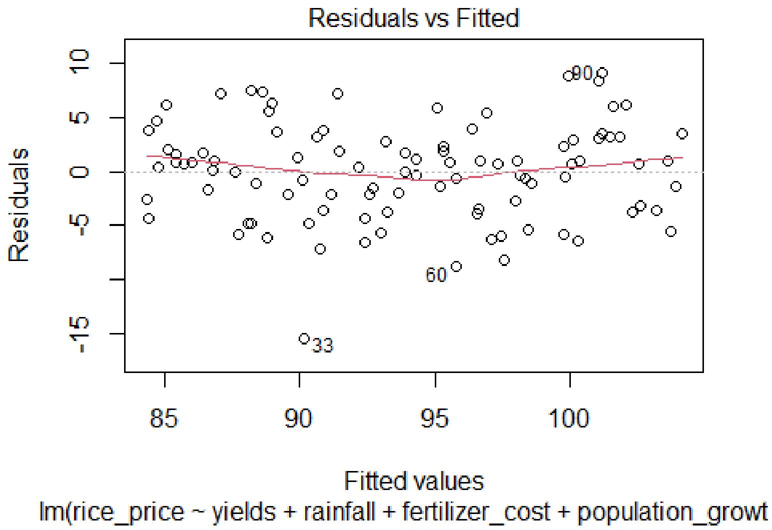

Figure 4.14: Residual vs fitted

Overall Pattern: The points in the scatter plot are randomly scattered around the horizontal line at zero, with no obvious patterns or trends. This is a good sign, as it suggests that the residuals are independent of the fitted values and that the model's assumptions are met.

Spread of the residuals: The residuals appear to be evenly spread out above and below the zero line, without any major outliers. This suggests that the model's variance is constant across the range of fitted values.

No strong bias: The residuals are centered around the zero line, indicating that the model is not consistently underestimating or overestimating the rice price.

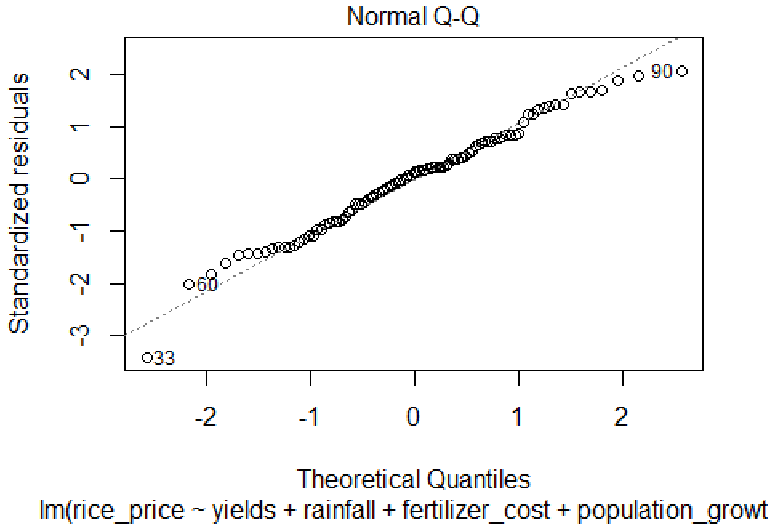

Figure 4.15: Normal Q-Q plot

Model fit: The adjusted R-squared of 0.6144 indicates that the model explains about 61.44% of the variation in rice price. This is a moderately high value, suggesting that the model is an approximately good fit for the data.

Predictor significance: The p-value for fertilizer cost is less than 0.05, indicating that this is a statistically significant predictor of rice price. Yields, rainfall, and population growth have p-values greater than 0.05, so their relationships with rice prices are not statistically significant based on this model.

Predictor direction and magnitude: Fertilizer cost and population growth have positive coefficients, meaning they are associated with higher predicted rice prices. Yields and rainfall have negative coefficients, suggesting they are associated with lower predicted rice prices. The absolute value of the fertilizer cost coefficient is the largest, indicating that it has the strongest impact on the predicted rice price among the four predictors.

Normality of residuals:

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.98327, p-value = 0.2426
```

Figure 4.16: Normality of residuals

The p-value of 0.2426 is greater than 0.05, so we cannot reject the null hypothesis. This means that there is not enough evidence to conclude that the residuals are not normally distributed.

Homoscedasticity:

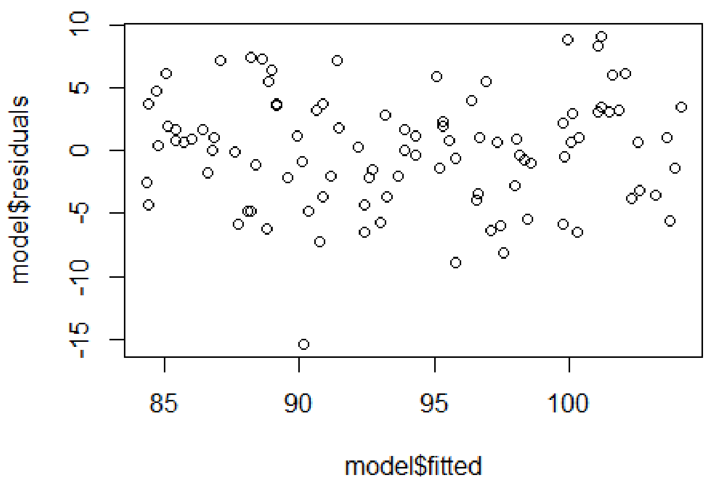

Figure 4.17: Homoscedasticity

The residuals are randomly scattered around the horizontal line at zero, with no obvious patterns or trends. This is a good sign, as it suggests that the residuals are independent of the fitted values and that the homoscedasticity assumption is met.

To create the best regression model, we followed the below step-wise methods by using R.

1. Forward direction
2. Backward Direction
3. Both Direction

```
summary(best_model)

##
## Call:
## lm(formula = rice_price ~ yields + rainfall + fertilizer_cost +
##     population_growth, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.3930  -3.2807   0.6043   3.0781   9.0749
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -9.243e+02  2.024e+02  -4.566 1.50e-05 ***
## yields            -6.113e-02  6.306e-02  -0.969    0.335
## rainfall          -7.171e-03  9.659e-02  -0.074    0.941
## fertilizer_cost    2.021e+01  4.148e+00   4.872 4.47e-06 ***
## population_growth  6.089e-01  1.051e+00   0.579    0.564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.537 on 94 degrees of freedom
## Multiple R-squared:  0.6301, Adjusted R-squared:  0.6144
## F-statistic: 40.04 on 4 and 94 DF,  p-value: < 2.2e-16
```

Figure 4.18: Summary of Forward Step wise model

According to the above figures, we can see that the Adjusted R-squared of 61.4%. It is the same adjusted R-squared of the initial model with all the independent variables without concerning their significance. Therefore the forward step-wise model has been selected for further analysis.

4.3 Time Series analysis

To increase the accuracy of the regression model, Time series effect was added to the initial data set.

```
##
## Call:
## lm(formula = rice_price ~ yields + rainfall + fertilizer_cost + population_growth
##      + data3$year + data3$month + data3$day, data = data3)
##
## Coefficients:
##       (Intercept)            yields          rainfall    fertilizer_cost
##        -2.773e+06        -4.859e-02        -3.296e-03         -3.571e+02
## population_growth        data3$year       data3$month          data3$day
##         4.468e-01         1.379e+03         1.147e+02          3.898e+00
```

Including time series analysis and the year has demonstrably improved the model's accuracy in explaining rice price variations. This signifies the importance of considering temporal trends in such analyses. The year's significant coefficient implies that rice prices might have a systematic dependence on time, which could be further investigated. This could involve exploring year-specific trends or incorporating more granular time units (e.g., month, season) as predictors.

4.3.1 Check the stationary of the new data set

```
adf.test (data3$rice_price)

##
##  Augmented Dickey-Fuller Test
##
## data:  data3$rice_price
## Dickey-Fuller = -3.7108, Lag order = 4, p-value = 0.02696
## alternative hypothesis: stationary
```

Figure 4.19: Augmented Dickey-Fuller Test

The test statistic and p-value come out to be equal to -3.7108 and 0.0296 respectively. Since the p-value is equal to or lesser than 0.05, hence we would reject the null hypothesis. It implies that the time series is stationary.

```
## Series: data3$rice_price
## ARIMA(0,1,1) with drift
##
## Coefficients:
##           ma1   drift
##       -0.9142  0.1982
## s.e.   0.0935  0.0445
##
## sigma^2 = 20.89:  log likelihood = -287.88
## AIC=581.76   AICc=582.01   BIC=589.51
```

Figure 4.20: Build the ARIMA Model

```
##
## Call:
## lm(formula = rice_price ~ yields + rainfall + fertilizer_cost +
##     population_growth + arima_residuals, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.92828 -0.95527  0.01643  0.90741  2.38800
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -858.38413   51.85653 -16.553   <2e-16 ***
## yields               0.01315    0.01627   0.808    0.421
## rainfall             0.01355    0.02474   0.548    0.585
## fertilizer_cost     18.81196    1.06259  17.704   <2e-16 ***
## population_growth    0.39043    0.26920   1.450    0.150
## arima_residuals      0.95870    0.02618  36.620   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.161 on 93 degrees of freedom
## Multiple R-squared:  0.976,   Adjusted R-squared:  0.9747
## F-statistic: 756.8 on 5 and 93 DF,  p-value: < 2.2e-16
```

Figure 4.21: Summary of the Final Regression model

At the end of our analysis, the final model has obtained an Adjusted R-squared value of 97.47%. Some of the independent variables in the model are statistically significant.

Chapter 5

Discussion and conclusions

5.1 Discussion

Our objective was to find a regression model for the rice data set and find an optimum price using our observed equation. We couldn't use the time effect for the regression model. Without the time effect, we got the equation and the accuracy of the equation was a little bit low. We check the accuracy of the model using the R-squared value. The observed value of the Adjusted R-squired value was less than eighty percent, that's why we conclude that the accuracy of the model is low.

To increase accuracy, we did the time series analysis for the price variable and the time. We fitted an ARIMA model first. After that, we get the error of the ARIMA model. The error of the ARIMA model is equal to the error of a Regression model. Using this method we get a new model and check the R-squired value. The value was very high and therefore we conclude that the accuracy of the model is very high.

5.2 Conclusion

The ordinary least squares method was used to build a multiple linear regression model to forecast the price of Rice depending on several variables such as yields, fertilizer cost, rainfall, population growth, and premium rice demand. All these variables have an effect on the rice price. When small changes in a variable fertilizer cost it will huge change of the Rice. The change in other variables has not a huge effect on price but they have a positive effect on Rice price.

Appendix

Extract of the dataset

The first 6 observations of rice prices along with the considered factors have been extracted from the data set.

rice price	yields	fertilizer cost
81.734	3817.87	50.01
88.109	3577.41	50.02
80.069	4596.26	50.03
87.086	3211.71	50.04
91.235	3451.069	50.05
85.144	4605.218	50.06

rice price	rainfall	population growth	premium rice demand
81.734	77.417	1.61781	-8.1552
88.109	70.074	1.0755	28.7423
80.069	71.817	1.76736	25.564
87.086	66.459	1.14968	6.0008
91.235	71.122	1.092	29.169
85.144	65.71	1.327	26.345

Bibliography

[1] Mumtaz Ali, Ravinesh C Deo, Nathan J Downs, and Tek Maraseni. Cotton yield prediction with markov chain monte carlo-based simulation model integrated with genetic programing algorithm: A new hybrid copula-driven approach. *Agricultural and forest meteorology*, 263:428–448, 2018.

[2] M Hossain. Economic prosperity in asia: implications for rice research. pages 3–16, 1996.

[3] Fernando Henrique Lermen, Márcia de Fátima Morais, Camila Matos, Rodrigo Röder, and Celise Röder. Optimization of times and costs of project of horizontal laminator production using pert/cpm technical. *Independent journal of management & production*, 7(3):833–853, 2016.

[4] Shashvi Mishra and Amit Kumar Tyagi. The role of machine learning techniques in internet of things-based cloud applications. *Artificial intelligence-based internet of things systems*, pages 105–135, 2022.

[5] Bibhuti B Mohanty. Regional disparity in agricultural development of maharashtra. *Economic and political weekly*, pages 63–69, 2009.

[6] Thong Nguyen-Huy, Ravinesh C Deo, Shahbaz Mushtaq, Jarrod Kath, and Shahjahan Khan. Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management. *Weather and climate extremes*, 21:76–89, 2018.

Optimal Pricing Using Multiple Regression Analysis: `https://www.routledge.com/Multiple-Regression-and-Beyond-An-Introduction-to-Multiple-Regression-and/Keith/p/book/9781138061446`

Time Series Analysis for Optimal Inventory Control: `https://www.sciencedirect.com/topics/neuroscience/time-series-analysis`