UiT Norges arktiske universitet

Fakultet for biovitenskap, fiskeri og økonomi

Indonesia rice farms

CASE 1

Kandidatnummer 33, 34 og 16

SOK-3011, Høst 2024

1 Introduction

In this assignment, we're going to find a dataset to use in solving the questions asked. We're going to collect data for estimating production functions, and then use them to estimate either a cost or profit function. At last, we will discuss and make a recommendation for which of the production functions that is the best fit for the data.

The dataset we are going to use is the RiceFarms dataset, found at the vincentralbundock github page, which is a dataset over different rice farms in Indonesia. Firstly, in chapter 2, we will download and load the dataset in Rstudio and then make a descriptive statistics table to show the included variables for our analysis. Then in chapter 3, we're doing three different production functions for the collected data to then test in a goodness-of-fit analysis. Further in chapter 4 we will estimate either a cost or profit function using a Cobb-Douglas specification, and then interpret our results. In the last in chapter 5, we will discuss and conclude our work.

2 The Data

In this paper, we're going to use the dataset, RiceFarms from the vincentralbundock github page. It is based on the research article by Qu Feng and William C. Horace with the title "Alternative technical efficiency measures: Skew, bias and scale". The authors say that this dataset has been analyzed a number of times (Feng & Horrace, 2012).

The dataset was collected by the Agro Economic Survey as a part of the Rural Dynamic Study of the rice production in Cimanuk River Basin and West Java. It was obtained by the Center for Agro Economic Research, Ministry of Agriculture, Indonesia (Erwidodo, 1990). The project surveyed 171 firms and was supposed to last for many years, but it concluded after two years with four total crops, due to a lack of funding. The project was reinstalled later to cover the wet season of 1982 to 1983 and the dry season of 1983 (Erwidodo, 1990).

2.1 Variables

The dataset contains 1026 observations of 21 variables. We will now present a short description of each of them, based on the information from the source document: [R: Production of Rice in Indonesia (vincentarelbundock.github.io)] we have the following variables,

rownames is just the number of the row and is not useful for this assignment.

id is the personal identification of a farm, each farm has a number that is identifying them.

size is the size of the production area the farm uses, measured in hectares.

status describes if it is a owner(includes lessee of the land), a shareholder or a mix of owner and shareholder.

varieties describes if it is one of 'trad' (traditional varieties), 'high' (high yielding varieties) or 'mixed' (mixed varieties) the farms are using.

bimas describes if the farm are participating in the biMAS intensification program, yes, no and mixed(means that not all of the land is registered to be in the program).

seed is how much seed a farm uses in the production measured in kilograms.

urea is a substance some farmers use as a sort of fertilizer (Uggerud , 2021) in the dataset it is measured in kilograms.

phosphate is a substance used as fertilizer in the production and is measured in kilograms.

pesticide is cost of pesticide in Rupiah

pseed is the kilograms price of seed in Rupiah

purea is the kilogram price of urea in Rupiah

pphosp is the price for phosphate in Rupiah per kilogram.

hiredlabor are the labor hired in measured in hours.

famlabor are the labor of family measured in hours.

totlabor is the total labor used for farming, both family labor and hired labor.

wage is the price of labor per hour in Rupiah.

goutput is the gross output of rice produced.

noutput is the net output of rice, it is the gross output minus the cost for harvesting(harvesting cost are paid in rice)

price is the pris for the rice per kg in Rupia.

region is the region in Indonesia the farm are located, the regions are, “warabinangun”, “langan”,“gunungwangi”, “malausma”,“sukaambit”, “ciwangi”.

with several variables we can calculate the input cost and the income and find the profit.

2.2 Descriptive statistics

If we see in table 1 the first descriptive statistics of the variables with number value, we have the mean, median, the lowest and the highest value and the first and the third quartiles.

Table 1:

Deskriptiv statistikk

size	seed	urea	phosphate	pesticide	pseed	purea	pphosph	hiredlabor	famlabor	totlabor	wage	goutput	noutput	price
Min. :0.0100	Min. : 1.00	Min. : 1.00	Min. : 0.00	Min. : 0	Min. : 40.0	Min. : 50.00	Min. : 60.00	Min. : 1	Min. : 1.0	Min. : 17.0	Min. : 30.00	Min. : 42.0	Min. : 42	Min. : 50.00
1st Qu.:0.1430	1st Qu.: 5.00	1st Qu.: 25.00	1st Qu.: 8.00	1st Qu.: 0	1st Qu.: 70.0	1st Qu.: 70.00	1st Qu.: 70.00	1st Qu.: 36	1st Qu.: 69.0	1st Qu.: 144.0	1st Qu.: 49.38	1st Qu.: 420.0	1st Qu.: 380	1st Qu.: 60.50
Median :0.2860	Median : 10.00	Median : 60.00	Median : 20.00	Median : 0	Median : 81.0	Median : 80.00	Median : 80.00	Median : 112	Median : 111.0	Median : 252.0	Median : 57.14	Median : 886.5	Median : 800	Median : 75.00
Mean :0.4316	Mean : 18.21	Mean : 95.44	Mean : 33.73	Mean : 595	Mean :112.1	Mean : 78.98	Mean : 79.57	Mean : 237	Mean : 151.5	Mean : 388.4	Mean : 80.42	Mean : 1405.2	Mean : 1241	Mean : 90.96
3rd Qu.:0.5000	3rd Qu.: 20.00	3rd Qu.: 100.00	3rd Qu.: 50.00	3rd Qu.: 265	3rd Qu.:150.0	3rd Qu.: 85.00	3rd Qu.: 85.00	3rd Qu.: 260	3rd Qu.: 185.0	3rd Qu.: 435.0	3rd Qu.:128.75	3rd Qu.: 1606.0	3rd Qu.: 1444	3rd Qu.:120.00
Max. :5.3220	Max. :1250.00	Max. :1250.00	Max. :700.00	Max. :62600	Max. :375.0	Max. :100.00	Max. :120.00	Max. :4536	Max. :1526.0	Max. :4774.0	Max. :175.35	Max. :20960.0	Max. :17610	Max. :190.00

We can see for example that the lowest amount of seed one farm is using is one kilogram, and the highest amount of seed used on a farm is 1250 kilograms.

In table 3 are we counting up all the observations we have in character variables so that we can see for example with the variable “status” that there are three categories and how many observations for which observation. Now since we have 6 observations per farm, we have not calculated that in this table, some of the farms may have different status over time, and it affects the quality of the counting.

Table 3

Counting of variables with character		
Variable	**Value**	**n**
bimas	mixed	162
bimas	no	779
bimas	yes	85
region	ciwangi	216
region	gunungwangi	222
region	langan	144
region	malausma	198
region	sukaambit	132
region	wargabinangun	114
status	mixed	211
status	owner	736
status	share	79
varieties	high	294
varieties	mixed	50
varieties	trad	682

3. Production functions

In this part we are going to analyse three different kinds of production functions while using the ricefarm dataset. We are going to start with the Cobb-Douglas function and work our way downward. The method used in chapter 3 and 4 is based on the works of Henningsen (2014).

3.1 Cobb-Douglas

We are going to set up a cobb-douglas function using our dataset ricefarm. A cobb douglas production function is a common approach to represents relationship between inputs as seed, urea, labor, size and phosphate and what level of production. The function is shown in the equation 3.1.

$$3.1\ \ln(\text{goutput}) = \beta_0 + \beta_1 \ln(\text{seed}) + \beta_2 \ln(\text{urea}) + \beta_3 \ln(\text{totlabor}) + \beta_4 \ln(\text{size}) + \beta_5 \ln(\text{phosphate}) + \epsilon$$

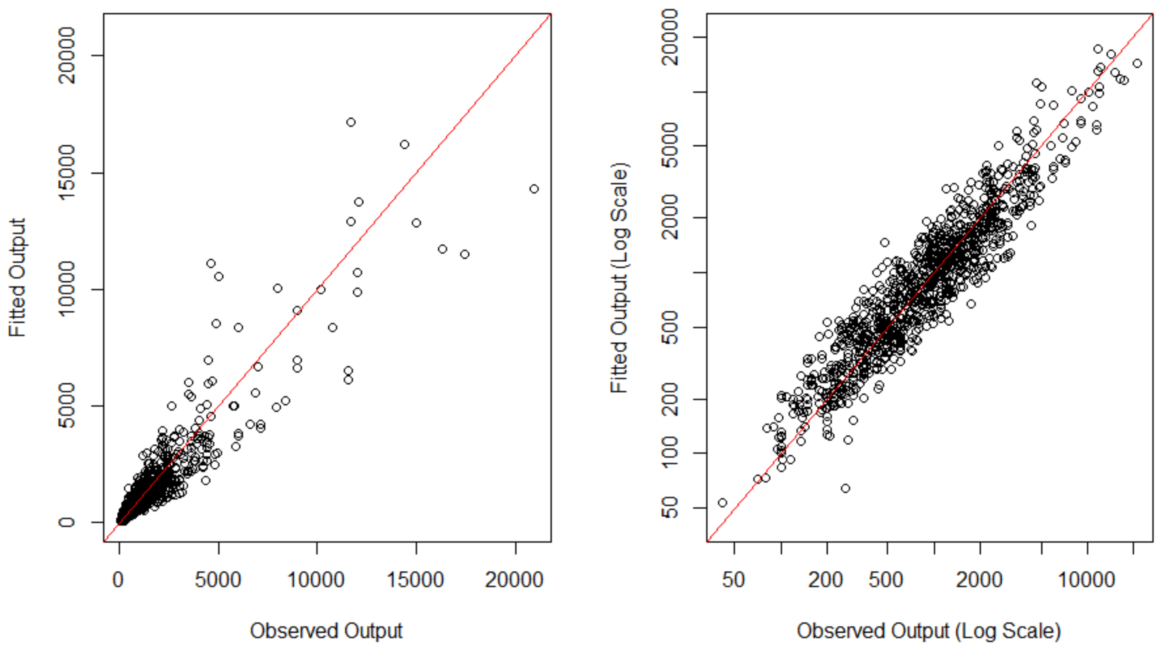
Where β_0 represents the constant, and the different β coefficients give us the different elasticities.

Figure 3.2

```
Call:
lm(formula = log_goutput ~ log_seed + log_urea + log_totlabor +
    log_size + log_phosphate, data = logmodel)

Residuals:
     Min       1Q   Median       3Q      Max
-1.12506 -0.23621  0.02195  0.23029  1.39828

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.033516   0.191154  26.332  < 2e-16 ***
log_seed      0.168172   0.026120   6.438 1.86e-10 ***
log_urea      0.139472   0.017091   8.160 9.77e-16 ***
log_totlabor  0.211613   0.028474   7.432 2.26e-13 ***
log_size      0.463683   0.030565  15.171  < 2e-16 ***
log_phosphate 0.063438   0.009856   6.437 1.88e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3508 on 1020 degrees of freedom
Multiple R-squared:  0.8762,	Adjusted R-squared:  0.8756
F-statistic:  1444 on 5 and 1020 DF,  p-value: < 2.2e-16
```

Our results are shown in the figure 3.2. We use cobb-douglas because it helps us interpret the coefficient as elasticities. Intercept: the intercept is 5.03. This means if all input were at a base level, the base level of output would be 5.03. Seed: 1% increase in the amount of seed would increase the rice output by 0.168%. p-value is less than 0.01. Effect is therefore both positive and significant. Urea fertilizer: 1 % increase in the amount of urea fertilizer would increase output by 0.139%. p-value < 0.01, therefore significant. Total labor: 1% increase in the amount of total labor, this includes both family labor and hired labor, would increase the rice output by 0.21%. p-value < 0.01, therefore significant. Size: 1 % increase in total land would increase rice output 0.46%. p-value < 0.01 therefore significant. This is the largest coefficient, which indicates that farm size has a big impact on production. Phosphate: 1% increase in phosphate would increase rice output by 0.06%. p-value < 0.01, therefore significant. R-squared is 0.876 which is excellent. 87.6% of the variation in rice production is explained by the inputs. The monotonicity condition is fulfilled, because all our coefficient inputs are non-negative. All these inputs increase production, which is the typical economic assumption.

Figure 3.3



3.1.1 Marginal products

Figure 3.4

```
    MP_seed          MP_urea          MP_totlabor        MP_size        MP_phosphate      MP_seed_obs        MP_urea_obs       MP_totlabor_obs   MP_size_obs       MP_phosphate_obs
 Min.   : 0.977   Min.   :  0.6949   Min.   :0.1507   Min.   : 495.3   Min.   : 0.316   Min.   : 0.5213   Min.   :  0.5021   Min.   :0.1908   Min.   :  185.5   Min.   : 0.370
 1st Qu.:10.790   1st Qu.:  1.5825   1st Qu.:0.5487   1st Qu.:1284.4   1st Qu.: 1.568   1st Qu.: 9.9835   1st Qu.:  1.5342   1st Qu.:0.4898   1st Qu.: 1104.1   1st Qu.: 1.507
 Median :13.923   Median :  1.9983   Median :0.6787   Median :1465.6   Median : 2.286   Median :13.4538   Median :  2.0874   Median :0.6727   Median : 1465.2   Median : 2.252
 Mean   :14.062   Mean   :  3.2827   Mean   :0.7292   Mean   :1470.8   Mean   : 2.847   Mean   :14.8828   Mean   :  4.0072   Mean   :0.7792   Mean   : 1568.1   Mean   : 2.893
 3rd Qu.:17.009   3rd Qu.:  2.6355   3rd Qu.:0.8635   3rd Qu.:1658.6   3rd Qu.: 3.298   3rd Qu.:18.4989   3rd Qu.:  2.8214   3rd Qu.:0.9552   3rd Qu.: 1867.6   3rd Qu.: 3.489
 Max.   :44.914   Max.   :180.3852   Max.   :2.1233   Max.   :5555.2   Max.   :32.612   Max.   :50.4516   Max.   :439.3379   Max.   :4.3243   Max.   :12751.3   Max.   :40.600
                                                                       NA's   :143                                                                            NA's   :143
```

What we have done in table 3.4 is to calculate the marginal products of the predicted values given a cobb-douglas production function versus the observed values. Marginal product measures the additional output when input is increased by one unit, holding all other input constant. The marginal product of the observed values measures how much additional output that actually was produced when a farm increased their inputs. Marginal product of the predicted output reflects the relationship estimated using our cobb-douglas model. These values will then remove the noise or the variability that is presented in the observed data. Noise is the fluctuations in the data that cannot be described by the model or the inputs. Our summary shows that a one unit increase in seed will on average increase output by around 14.06 or 14.8. If the firm increases urea by one unit output will increase on average 3.28 or or 4. One unit increase in labor increases output on average by 0.729 or 0.7792. One unit increase of phosphate increases output on average by 2.847 or 2.893. We can observe that the max values for size are very different, respectively observed values 12751 and predicted 5555. This obviously raises questions as to why that is. First of all, it is important to understand that a cobb-douglas model is based on simplifying assumptions. The model uses a log-linear relationship which might not be flexible enough to account for the extreme values captured in the observed data. Also, there might be measurement errors in the observed output that inflate the marginal product of size. There might be a few farms that have unusual high productivity due to favorable weather, unexpected market conditions, or even reporting errors. Therefore, the predicted marginal products might show a more accurate long-term reflection of the ricefarm market since it uses a more moderate estimate.

3.1.2 Return to scale

To calculate return to scale for all firms in a Cobb-Douglas production function, all we do is sum all the coefficients except the intercept. In rice farms if we increase all input by 1 unit, the output will increase by 1.046379. Hence, we have although small, increasing return to scale.

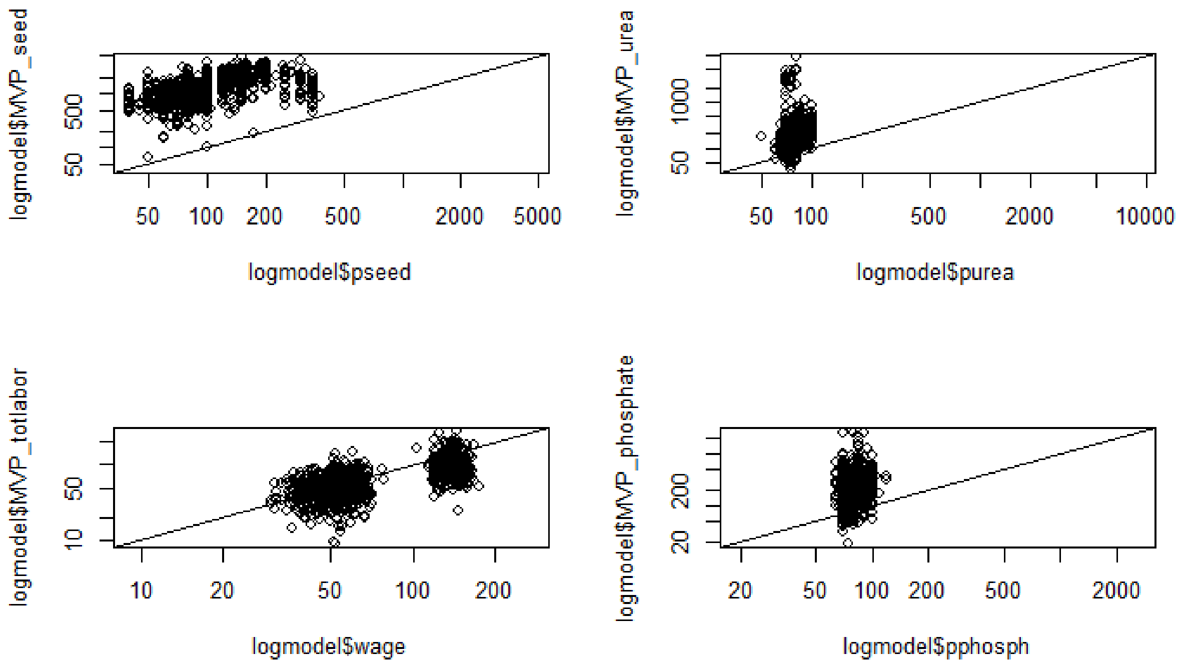
Figure 3.5

```
elasticity_of_scale <- sum(coef(cobb_douglas)[-1]
elasticity_of_scale

 [1] 1.046379
```

3.1.3 Profit Maximizing behavior

We can show the profit maximizing behavior as shown in figure 3.6. We plot marginal products against the input prices across firms, but we are excluding size since our dataset does not have a leasing price or a price per hectare. Figure 3.6 shows the marginal value products for most of them is higher than the corresponding input prices. Most of the rice firms could increase their profits by seed. Wage however centers around the middle, and even some firms have a decreasing return to scale. Urea and phosphate however most firms would increase rice production by increasing these inputs.

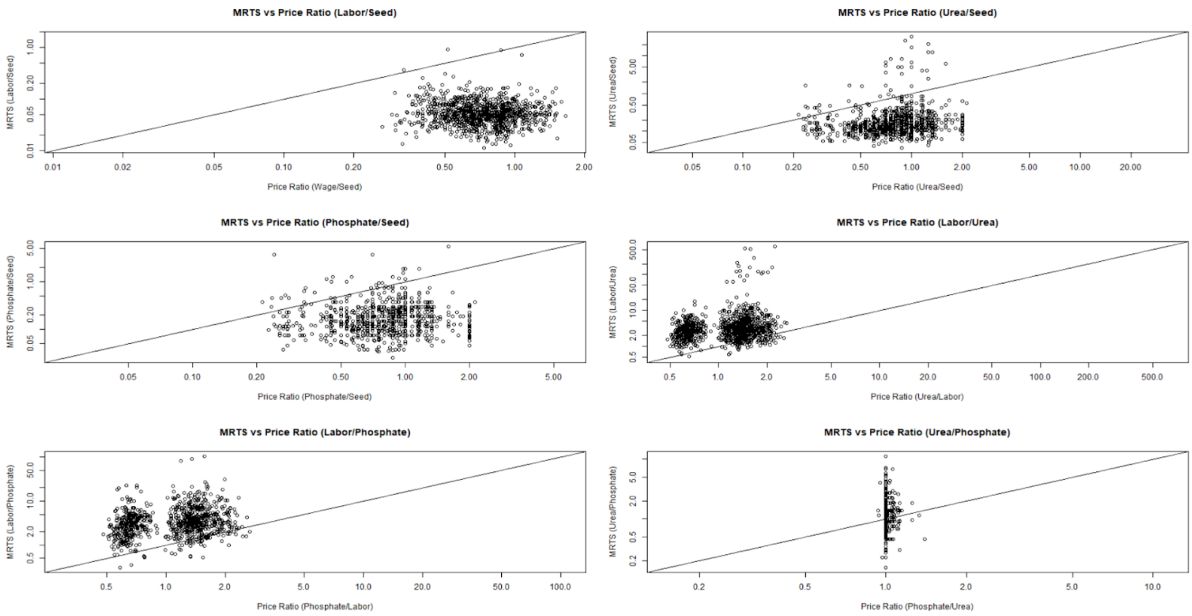
Figure 3.6



3.1.4 Cost minimization

Here we want to interpret the MRTS against the price ratio of input prices. Cost minimization principles suggest the marginal rate of technical substitution should equal the ratio of the input prices. As seen in the plots below most of these farms do not align with the cost minimization principle where MRTS = price ratio. We exclude size since our dataset does not have a price. We see that the MRTS for labor and seed is far lower than the price ratio for most farms. This indicates that farms may underuse labor relative to seed, based on the cost for labor. These plots reveal inefficiency among most farms, as there are quite big deviations from the price ratio line. This suggests room for improvement in efficiency especially labor, urea and phosphate as seen in Figure 3.7.

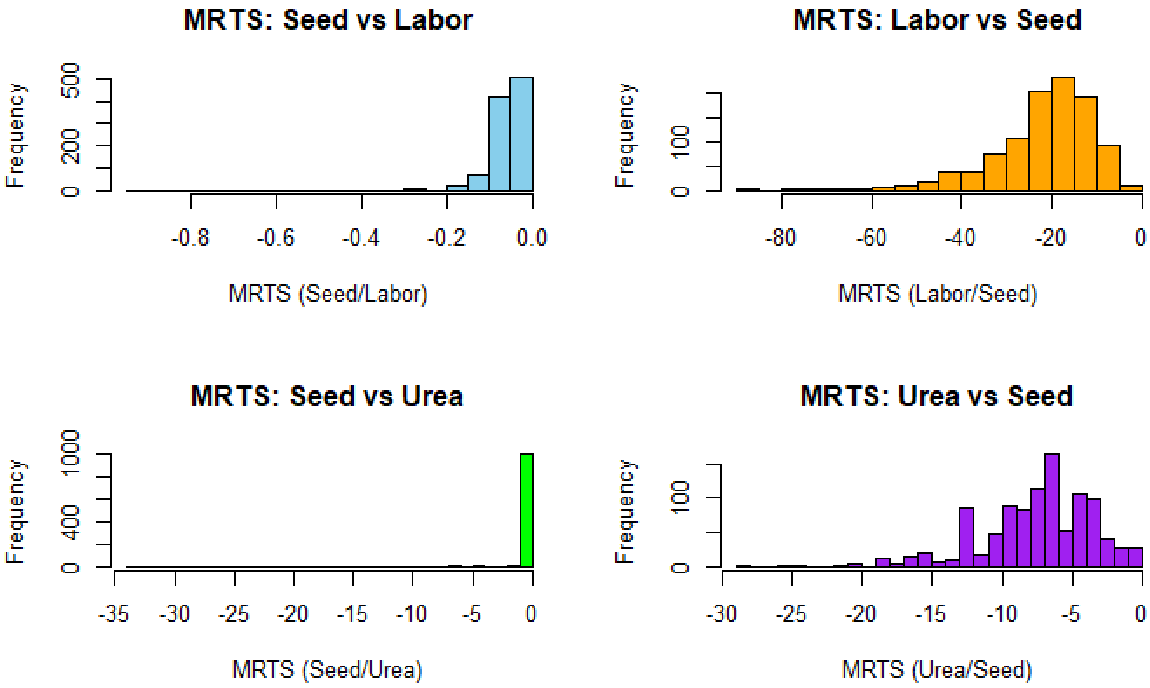
Figure 3.7



3.1.5 Marginal rates of technical substitution

Further we calculate each marginal rate of technical substitution for all input pairs. We have calculated MRTS between seed and labor, seed and urea, seed and size, seed and phosphate, labor and urea, labor and size and labor and phosphate, urea and size, urea and phosphate, size and phosphate. that represents each of the slopes isoquant curves, which is a curve showing all possible combinations of inputs that result in the same level of output.

Figure 3.8



Each of the plots in figure 3.8 represents the rate at which farms can substitute different inputs while maintaining the same level of output. Example labor vs seed would therefore mean at what rate a farm can substitute labor for seed while maintaining the same level of output. All values are negative which is expected since it is a cobb-douglas function and inputs are not perfectly interchangeable. Some key observations in our MRTS are seed and labor, they are relatively symmetrical and concentrated around a small range, indicating a balanced trade-off. Certain input pairs such as labor vs phosphate, phosphate vs labor and urea vs phosphate exhibit zero values, this means that some farms did not use or record these input values. Some farms recorded zero phosphate usage, resulting in impossible to calculate MRTS. Instead of removing these values we have opted to show them, reflecting the different production strategies different ricefarm have. Some of the plots is not shown but is in the appendix.

3.1.6 Relative marginal rates of technical substitution

Figure 3.9

RMRTS_Seed_Labor <dbl>	RMRTS_Seed_Urea <dbl>	RMRTS_Seed_Size <dbl>	RMRTS_Seed_Phos... <dbl>	RMRTS_Labor_Urea <dbl>	RMRTS_Labor_Size <dbl>	RMRTS_Labor_Pho... <dbl>	RMRTS_Urea_Size <dbl>
0.794715	1.205773	0.3626871	2.65097	1.51724	0.4563738	3.33575	0.3007922

RMRTS_Urea_Phosph... <dbl>	RMRTS_Size_Phosphate <dbl>
2.198565	7.309249

The same with relative marginal rates of technical substitution, or RMRTS for short. The main difference between RMRTS and MRTS is the unit measurements. RMRTS is ratio elasticities. Interpretation of our data is that for every 1 percent decrease in seed, you must increase labor by 0.794% to get the same output. There is some substitutability between seed and labor, but the rate is relatively low. 1 percent decrease in seed, the ricefarm has to increase urea by 1.2% for the same output, and so on. Size and phosphate have high value. Interpretation of this is if you decrease size by 1%, you will have to increase phosphate by 7.3% to get the same output as before. Reason for this probably lies in the fact that size has a big effect on a farm's productivity, while some farms do not even use phosphate.

3.2 Quadratic production function

We will use the quadratic production function because the model, which is similar to the linear model, uses quadratic and interaction terms as well. This can be useful to calculate the actual production of the farms in our dataset.

Figure 4

```
Call:
lm(formula = goutput ~ seed + urea + totlabor + size + phosphate +
    I(0.5 * seed^2) + I(0.5 * urea^2) + I(0.5 * totlabor^2) +
    I(0.5 * size^2) + I(0.5 * phosphate^2) + I(seed * urea) +
    I(seed * totlabor) + I(seed * size) + I(seed * phosphate) +
    I(urea * totlabor) + I(urea * size) + I(urea * phosphate) +
    I(totlabor * size) + I(totlabor * phosphate) + I(size * phosphate),
    data = df_rice)

Residuals:
    Min      1Q  Median      3Q     Max
-5863.3  -211.2    -0.5   157.8  4986.6

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -6.061e+01  3.700e+01  -1.638 0.101740
seed                    9.673e+00  3.308e+00   2.924 0.003532 **
urea                    2.484e+00  6.863e-01   3.619 0.000311 ***
totlabor               -1.602e-01  1.928e-01  -0.831 0.406328
size                    2.110e+03  2.193e+02   9.626  < 2e-16 ***
phosphate               7.058e+00  1.111e+00   6.354 3.17e-10 ***
I(0.5 * seed^2)        -5.829e-03  2.211e-02  -0.264 0.792090
I(0.5 * urea^2)        -1.114e-02  4.209e-03  -2.646 0.008268 **
I(0.5 * totlabor^2)    -9.806e-04  4.664e-04  -2.102 0.035763 *
I(0.5 * size^2)        -2.904e+03  6.211e+02  -4.676 3.33e-06 ***
I(0.5 * phosphate^2)    2.574e-03  1.089e-02   0.236 0.813207
I(seed * urea)         -3.472e-03  1.284e-02  -0.270 0.786905
I(seed * totlabor)      1.572e-03  5.377e-03   0.292 0.770048
I(seed * size)         -2.217e+00  5.480e+00  -0.405 0.685888
I(seed * phosphate)     2.170e-02  4.005e-02   0.542 0.587968
I(urea * totlabor)     -4.328e-03  1.256e-03  -3.447 0.000591 ***
I(urea * size)          6.281e+00  1.458e+00   4.307 1.81e-05 ***
I(urea * phosphate)    -1.358e-02  5.938e-03  -2.288 0.022367 *
I(totlabor * size)      1.825e+00  4.472e-01   4.081 4.83e-05 ***
I(totlabor * phosphate) 8.875e-03  1.732e-03   5.124 3.59e-07 ***
I(size * phosphate)    -4.326e+00  2.151e+00  -2.011 0.044608 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 633.1 on 1005 degrees of freedom
Multiple R-squared:  0.8936,	Adjusted R-squared:  0.8915
F-statistic:   422 on 20 and 1005 DF,  p-value: < 2.2e-16
```

The quadratic production function in figure 4 shows a residual standard error of 633.1 on 1005 degrees of freedom, which measures how much the actual output varies from the expected output. The multiple R-squared is at 0.8936, which means that this model explains 89.36% of the variation in gross output. The adjusted R-squared is 0.8915, which is also high and means that the model is robust. The p-value is very small and highly significant, and therefore indicating that at least some predictors significantly explain the variation in gross output.

The estimates show the unit increase in gross output for adding one more unit of each input. The first 5 estimates are in linear terms and show way different estimates than the linear model does. The next 5 estimates are in quadratic terms. They seem to all show reduced returns for an increase in input, with statistical significance. Only phosphate has a positive estimate, but it isn't statistically significant. The last 10 estimates are in interaction terms, and show us how the combination of two inputs influence the output. Of this only urea with size, total labor with size and total labor with phosphate give a positive and statistically significant estimate. Urea with total labor, urea with phosphate and size with phosphate is showing negative and statistically significant estimates.

The quadratic model could be compared to the linear model because they are kind of similar. To test which model that best fit our dataset, we estimated a linear model to test by first using an F-test between the models as shown in figure 4.1

Figure 4.1

```
Wald test

Model 1: goutput ~ seed + urea + totlabor + size + phosphate
Model 2: goutput ~ seed + urea + totlabor + size + phosphate + I(0.5 *
    seed^2) + I(0.5 * urea^2) + I(0.5 * totlabor^2) + I(0.5 *
    size^2) + I(0.5 * phosphate^2) + I(seed * urea) + I(seed *
    totlabor) + I(seed * size) + I(seed * phosphate) + I(urea *
    totlabor) + I(urea * size) + I(urea * phosphate) + I(totlabor *
    size) + I(totlabor * phosphate) + I(size * phosphate)
  Res.Df Df      F    Pr(>F)
1   1020
2   1005 15 15.458 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If the F-value is high, and the p-value is low, then you can reject the null hypothesis, which is in this example that the linear model is sufficient for our data. As you can see in the model above, the quadratic model is a significantly better model for the data being used.

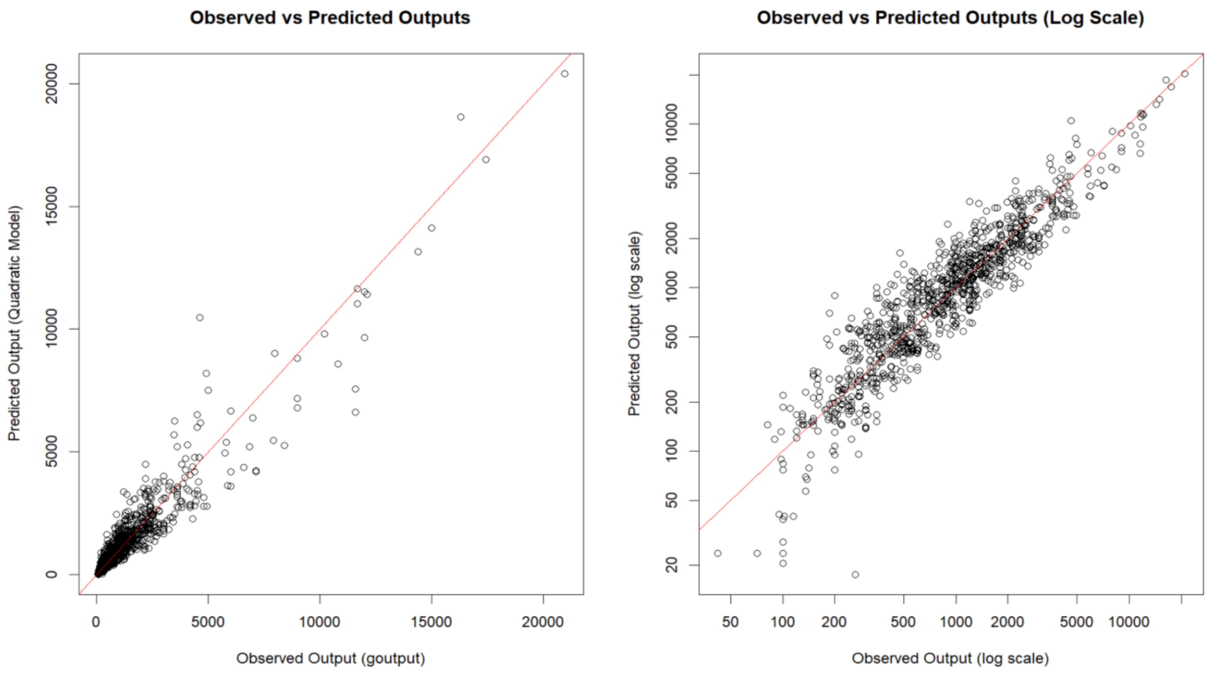
To be certain, we will check once more using a Chi squared test to determine if the quadratic model more likely to be a better fit than the linear model.

Figure 4.2

```
Likelihood ratio test

Model 1: goutput ~ seed + urea + totlabor + size + phosphate
Model 2: goutput ~ seed + urea + totlabor + size + phosphate + I(0.5 *
    seed^2) + I(0.5 * urea^2) + I(0.5 * totlabor^2) + I(0.5 *
    size^2) + I(0.5 * phosphate^2) + I(seed * urea) + I(seed *
    totlabor) + I(seed * size) + I(seed * phosphate) + I(urea *
    totlabor) + I(urea * size) + I(urea * phosphate) + I(totlabor *
    size) + I(totlabor * phosphate) + I(size * phosphate)
  #Df  LogLik Df Chisq Pr(>Chisq)
1   7 -8170.1
2  22 -8063.6 15   213  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The result shown in figure 4.2 is also a low p-value, which means the quadratic model is likely to be a better model for the data, and it is proven to be so twice.

Figure 4.3



The plots in figure 4.3 is to show the predicted values for the inputs and outputs for the rice farms used in the dataset. The plot to the left is quantified with a linear scale, and the plot to the right is quantified using a logarithmic scale. The reasoning behind showing the two different plots is because most of the farms are located close to 0,0 in the first plot. When using a logarithmic scale, you can easily see visually the difference between the farms, and the outliers especially at the lower parts showing.

To test whether this is useless, we will check to see if any of the predicted output quantities is false.

```
FALSE  TRUE
    1  1025
```

Unfortunately, one of the predicted outputs is false. To make completely sure that we can still use these plots, we checked the plots after removing the one false predicted output quantity. There is no visual difference, and data could therefore be considered quite robust. We will be keeping the plots and removing the false predicted output quantity for the rest of the analysis using the quadratic model.

Figure 4.4

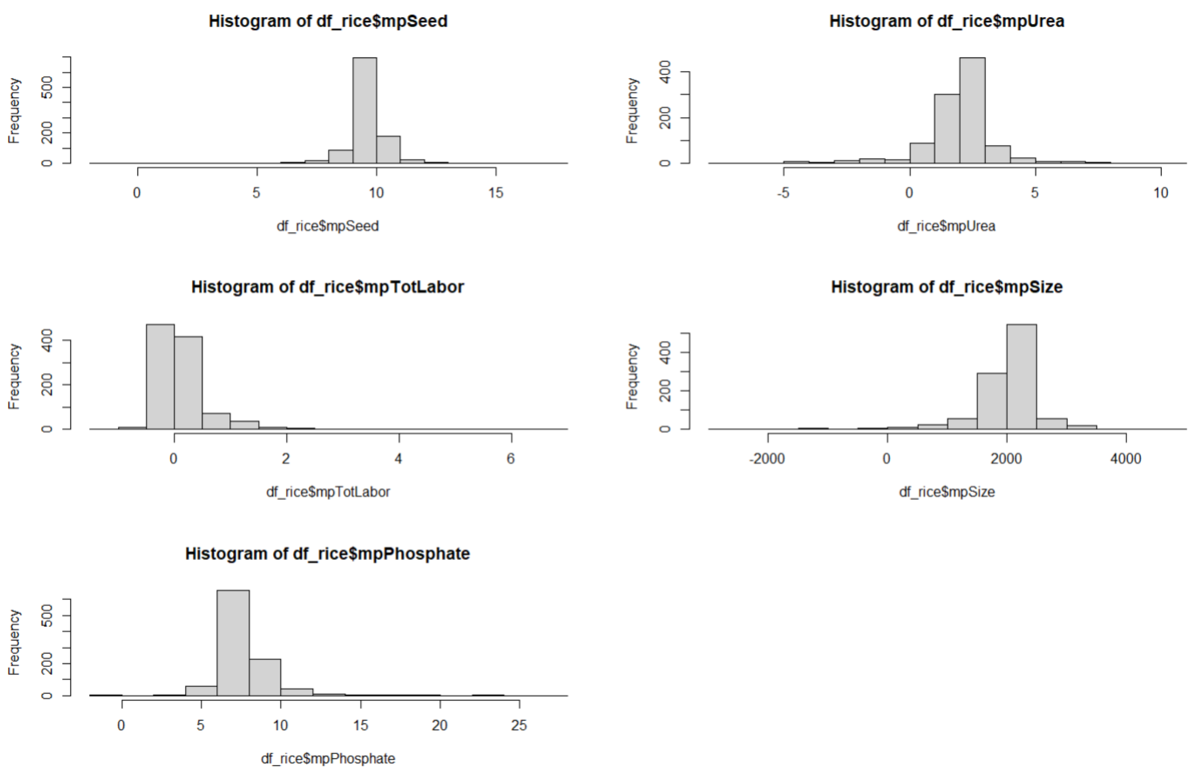


Figure 4.4 shows us how a unit increase for each input will increase output units for almost every farm. It shows that the monotonicity condition is not satisfactory for all observations, because they are not all equal to or greater than zero. If a farm increases seed input by one unit, most farms will increase output by around 9 units. If the input for urea increases by one unit, most farms output will increase by around 3 units. For one unit increase in labor, most farms won't see an increase or decrease in output unit. For size, the increase will be around 2400 units. For phosphate, around 7 units.

3.2.2 Elasticities quadratic model

Figure 4.5

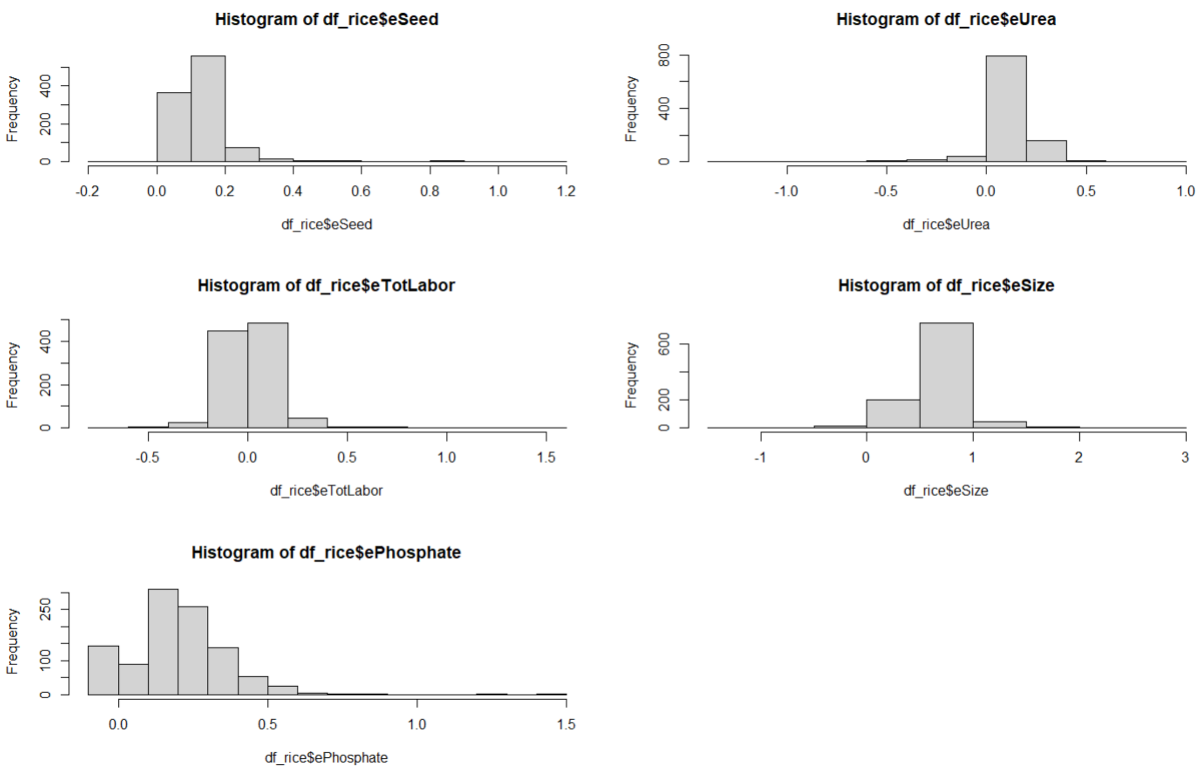


Figure 3.6 shows us that when you increase one of the input factors by one percent, the increase in output is typically positive for most farms. For seed, a one percent input increase will lead to around 0.12% increase in output. A one percent increase in urea will increase output by around 0.16% for most farms. For labor, the increase is around 0.06%. For size, the increase is around 0.8%. And for phosphate, the increase will be around 0.2%. There are some

cases where a one percent increase in urea, labor and phosphate, will decrease the farms output.

It can be helpful to visualize the elasticity for all input factors as one

Figure 4.6.

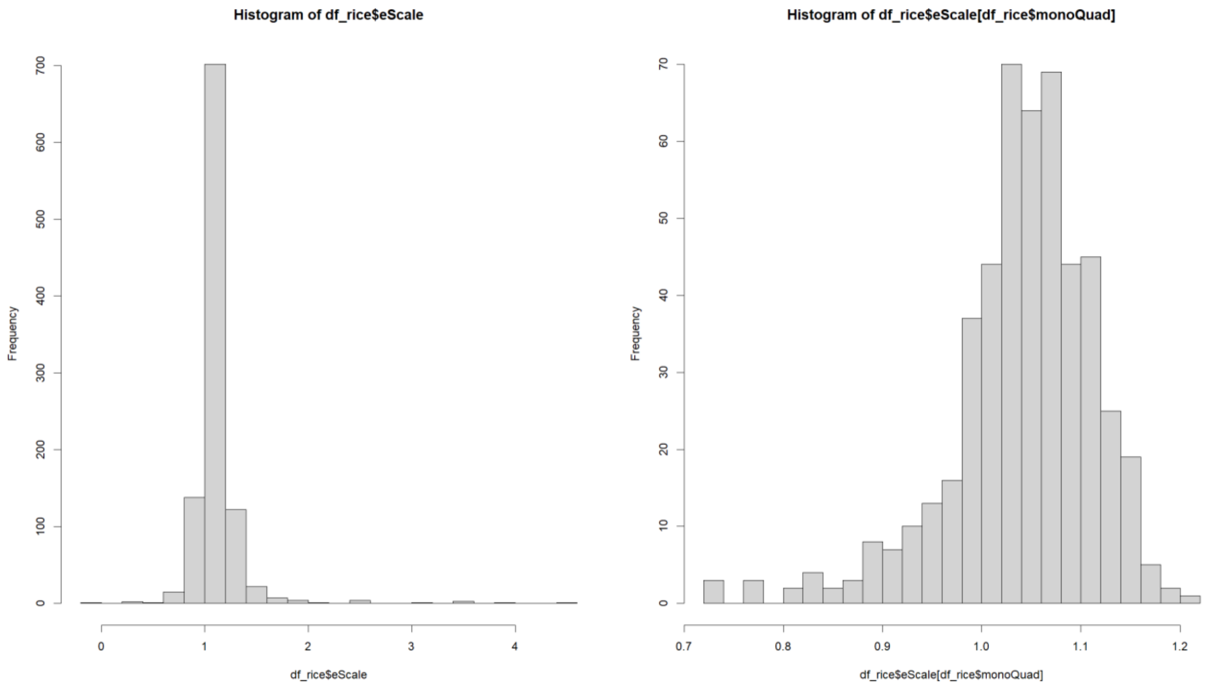
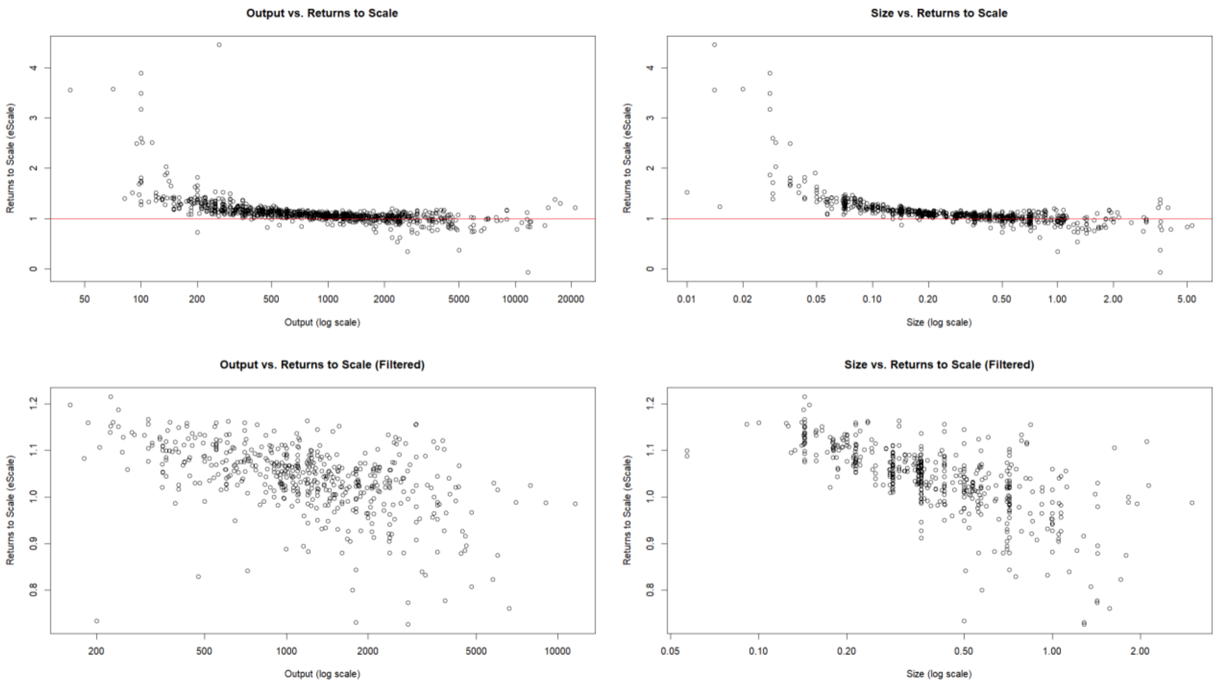


Figure 4.6 are histograms over the elasticities for each farm. As shown, most farms have an elasticity around 1.3 percent, with some outliers who have a lot higher elasticities. These histograms show us that increasing the inputs by one percent, will on average increase output by 1.3 percent, though some farms higher increasing return to scale.

Figure 4.7



As seen in figure 4.7, all farms would gain from increasing their input factor. The plots show us the optimal farm size when we analyse the relationship between the elasticity of scale and the farm size.

3.3 Translog production function

Further we are going to use the translog-model as it is more flexible than the cobb-douglas function, as it includes interaction effects and quadratic terms between inputs. In figure 4.8 we show you the model-

Figur 4.8

```
Call:
lm(formula = log_goutput ~ log_seed + log_urea + log_totlabor +
    log_size + log_phosphate + I(0.5 * log_seed^2) + I(0.5 *
    log_urea^2) + I(0.5 * log_totlabor^2) + I(0.5 * log_size^2) +
    I(0.5 * log_phosphate^2) + I(log_seed * log_urea) + I(log_seed *
    log_totlabor) + I(log_seed * log_size) + I(log_seed * log_phosphate) +
    I(log_urea * log_totlabor) + I(log_urea * log_size) + I(log_urea *
    log_phosphate) + I(log_totlabor * log_size) + I(log_totlabor *
    log_phosphate) + I(log_size * log_phosphate), data = logmodel)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0572 -0.2320  0.0305  0.2270  1.2848

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                3.1238160  1.8357166   1.702  0.08912 .
log_seed                   1.0823064  0.3763188   2.876  0.00411 **
log_urea                   0.1738504  0.2665779   0.652  0.51445
log_totlabor               0.6122342  0.5894316   1.039  0.29920
log_size                   0.1930834  0.4548992   0.424  0.67133
log_phosphate             -0.3941546  0.1658584  -2.376  0.01767 *
I(0.5 * log_seed^2)       -0.0612036  0.0495922  -1.234  0.21744
I(0.5 * log_urea^2)        0.0095123  0.0285470   0.333  0.73904
I(0.5 * log_totlabor^2)    0.0362934  0.1017729   0.357  0.72146
I(0.5 * log_size^2)        0.0486178  0.0686539   0.708  0.47901
I(0.5 * log_phosphate^2)   0.0600032  0.0148584   4.038 5.79e-05 ***
I(log_seed * log_urea)     0.0060822  0.0381387   0.159  0.87333
I(log_seed * log_totlabor) -0.1473921  0.0625038  -2.358  0.01856 *
I(log_seed * log_size)     0.1127936  0.0561727   2.008  0.04491 *
I(log_seed * log_phosphate) 0.0470229  0.0240509   1.955  0.05084 .
I(log_urea * log_totlabor) -0.0388989  0.0377993  -1.029  0.30368
I(log_urea * log_size)    -0.0684277  0.0455266  -1.503  0.13315
I(log_urea * log_phosphate) 0.0315920  0.0193205   1.635  0.10233
I(log_totlabor * log_size) 0.0997278  0.0721704   1.382  0.16733
I(log_totlabor * log_phosphate) -0.0009088  0.0246135  -0.037  0.97055
I(log_size * log_phosphate) -0.0862977  0.0275466  -3.133  0.00178 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3366 on 1005 degrees of freedom
Multiple R-squared:  0.8877,	Adjusted R-squared:  0.8855
F-statistic: 397.2 on 20 and 1005 DF,  p-value: < 2.2e-16
```

The p-value at seed is significant at 1% level, meaning 1% increase in seed should increase output by 1.08% roughly. The coefficients at the linear level are similar as done in the cobb douglas function, though they are not statistically significant. At the quadratic terms, eg 0.5*log_seed^2 etc are trying to capture the non-linear effects. A translog function also tries to capture the interaction between how one input changes when another input is changed. Three of the parameters are statistically significant at the 5% level. Seed and size have a coefficient at 0.114 with a p-value at 0.044. Meaning they have a positive interaction, meaning increasing both seed and size would increase productivity therefore increase output. Size and phosphate are significant but have a coefficient of -0.08, meaning increasing both would decrease output. R squared is high at 0.8877.

We are going to compare the cobb-douglas mode against the translog model to test whether the additional interaction input and quadratic terms in the translog model do significantly improve the model. For this we are using a WALD test in figure 4.9 which is simple to use in R. We set up our hypothesis like this:

Null hypotheses: the cobb Douglas model acceptable, meaning the translog model does not improve the model.

alternative hypotheses: The translog model provides a better fit, meaning it improves the overall model.

Figure 4.9

```
Wald test

Model 1: log_goutput ~ log_seed + log_urea + log_totlabor + log_size +
    log_phosphate
Model 2: log_goutput ~ log_seed + log_urea + log_totlabor + log_size +
    log_phosphate + I(0.5 * log_seed^2) + I(0.5 * log_urea^2) +
    I(0.5 * log_totlabor^2) + I(0.5 * log_size^2) + I(0.5 * log_phosphate^2) +
    I(log_seed * log_urea) + I(log_seed * log_totlabor) + I(log_seed *
    log_size) + I(log_seed * log_phosphate) + I(log_urea * log_totlabor) +
    I(log_urea * log_size) + I(log_urea * log_phosphate) + I(log_totlabor *
    log_size) + I(log_totlabor * log_phosphate) + I(log_size *
    log_phosphate)
  Res.Df Df      F    Pr(>F)
1   1020
2   1005 15 6.8717 2.246e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The wald test shows us a F-value of 6.8717 and an extremely low p-value, meaning that we can reject the null hypothesis. This means that the translog model provides a better fit, and the additional quadratic and interaction terms does improve the overall model.

Next, we perform a likelihood test, also called a LR-test. We use the same hypotheses as above, to see what model has a better goodness of fit, cobb-douglas or translog as shown in figure 5.

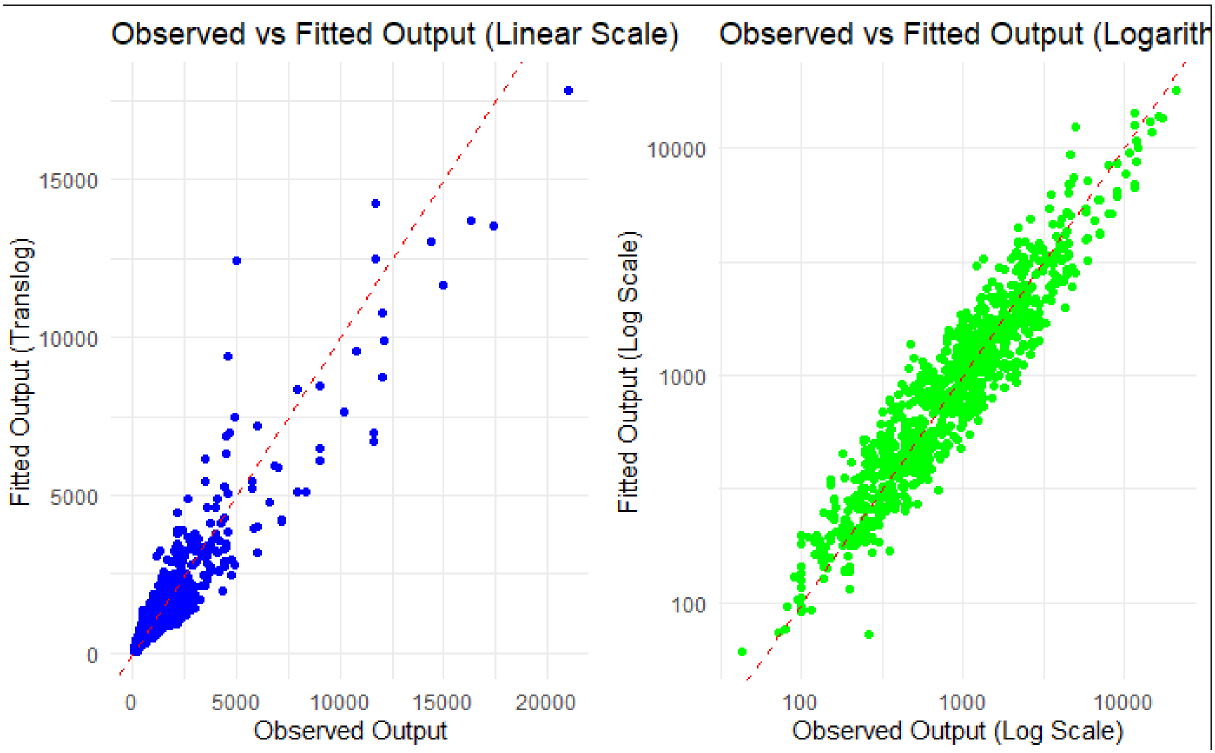
Figure 5

```
Likelihood ratio test

Model 1: log_goutput ~ log_seed + log_urea + log_totlabor + log_size +
    log_phosphate
Model 2: log_goutput ~ log_seed + log_urea + log_totlabor + log_size +
    log_phosphate + I(0.5 * log_seed^2) + I(0.5 * log_urea^2) +
    I(0.5 * log_totlabor^2) + I(0.5 * log_size^2) + I(0.5 * log_phosphate^2) +
    I(log_seed * log_urea) + I(log_seed * log_totlabor) + I(log_seed *
    log_size) + I(log_seed * log_phosphate) + I(log_urea * log_totlabor) +
    I(log_urea * log_size) + I(log_urea * log_phosphate) + I(log_totlabor *
    log_size) + I(log_totlabor * log_phosphate) + I(log_size *
    log_phosphate)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   7 -378.11
2  22 -328.02 15 100.18  1.208e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Log likelihood of cobb douglas is -378.11 and for translog its -328.08. It means that the log likelihood is higher in the translog model, which indicates that the translog model is a better fit. The chi-squared is 100.18, which is calculated computing the differences in the log likelihood of the two models. P-value is far less than 0.05, meaning that we reject the null hypothesis, and accept that the additional quadratic and interaction terms in the translog model do have a significantly better fit than the cobb-douglas model.

As we did before we want to compare the fitted model against the observed values. The left side uses linear scaling, where we can see absolute values, and on the right uses logarithmic scale. Points close to the red line indicates a good fit of the model, here the predicted values closely match the observed values. Both these plots showcase a relatively good fit of the model, there are obviously some outliers, especially in the linear scaling. Using the logarithmic scale is easier to visualize, since it takes to account small and large outputs on a proportional basis. Logarithmic scale deviates to a certain degree from the fitted line.

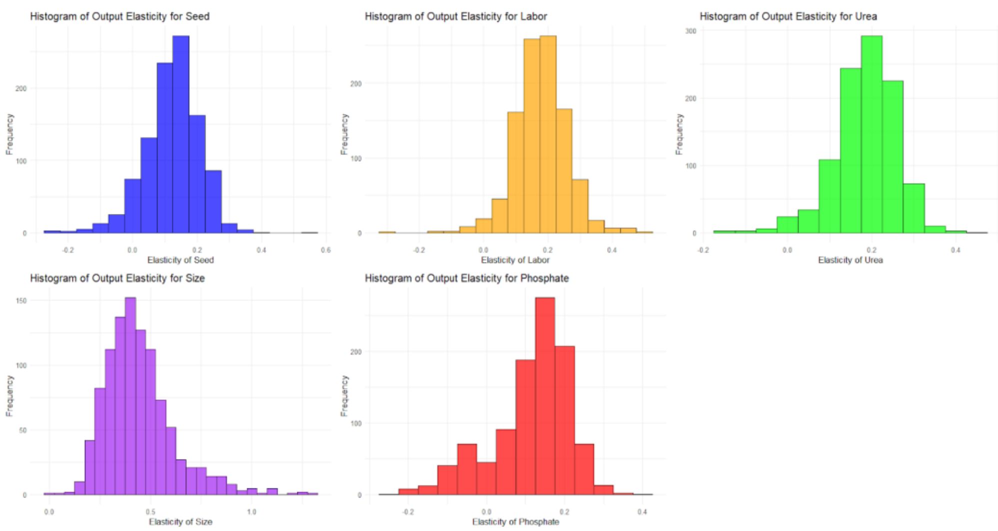
Figure 5.1



3.3.1 Output elasticities

Output elasticities measure percentage change in output based on a 1% change in input. We have done it for farms under and as we see an increase of 1% in labor average around 0.2% in rice production for farms. It's pretty similar output changes for seed, labor, urea and phosphate but for size however is definitely the biggest impact on rice production.

Figure 5.2



The output around size is centred around 0.5 %, which indicates that size plays a major role as seen on the other models for rice production as seen in figure 5.2. Since some of the observations hovers around minus change in output, it indicates to us that the monotonicity condition is not fulfilled. As we can see under some rice farms have a negative output elasticity of phosphate, as shown in appendix.

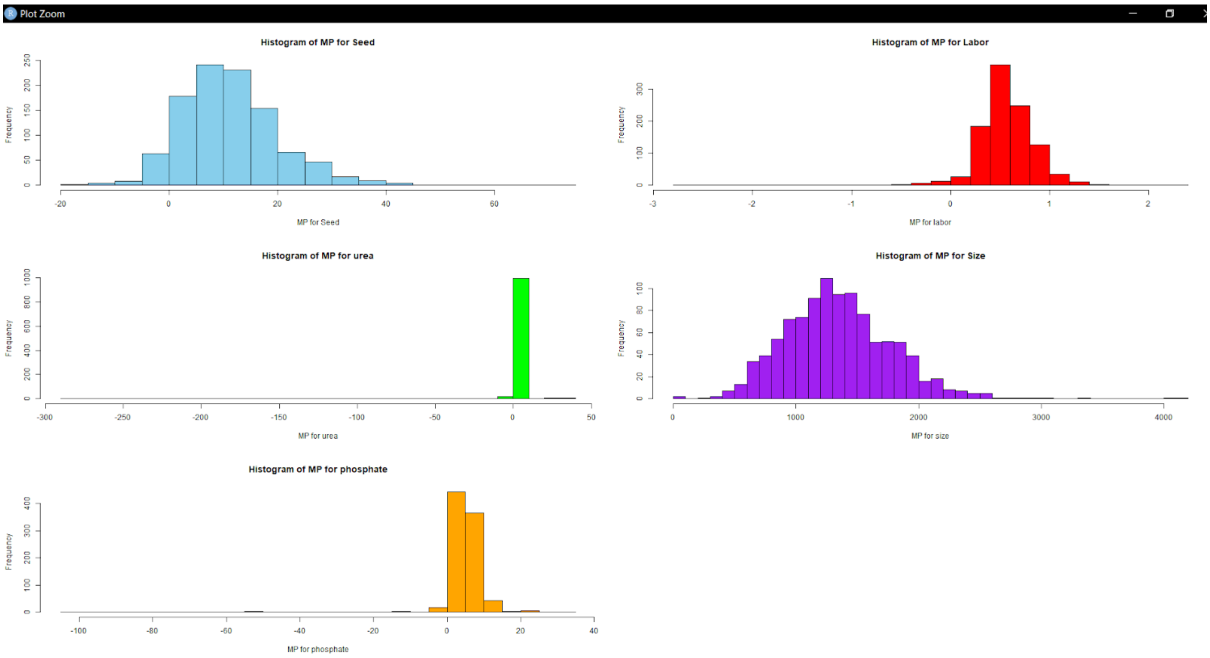
```
FALSE  TRUE
  193   833
```

193 farms do have a negative output elasticity of different inputs in production, therefore the monotonicity condition is not fulfilled.

3.3.2 Marginal product translog model

We see in figure 5.3 that the majority of the rice farms have a positive marginal product. Marginal product tries to highlight which input has the largest impact on rice production. In our model we can see that size is centred around 1400 units, which means that one unit increase in size will increase rice production around 1400 units. Some of the observations deviate from the mean, which leads to having negative effects of productivity This means that some of the farms may be inefficient with the use of seed, labor, urea and phosphate.

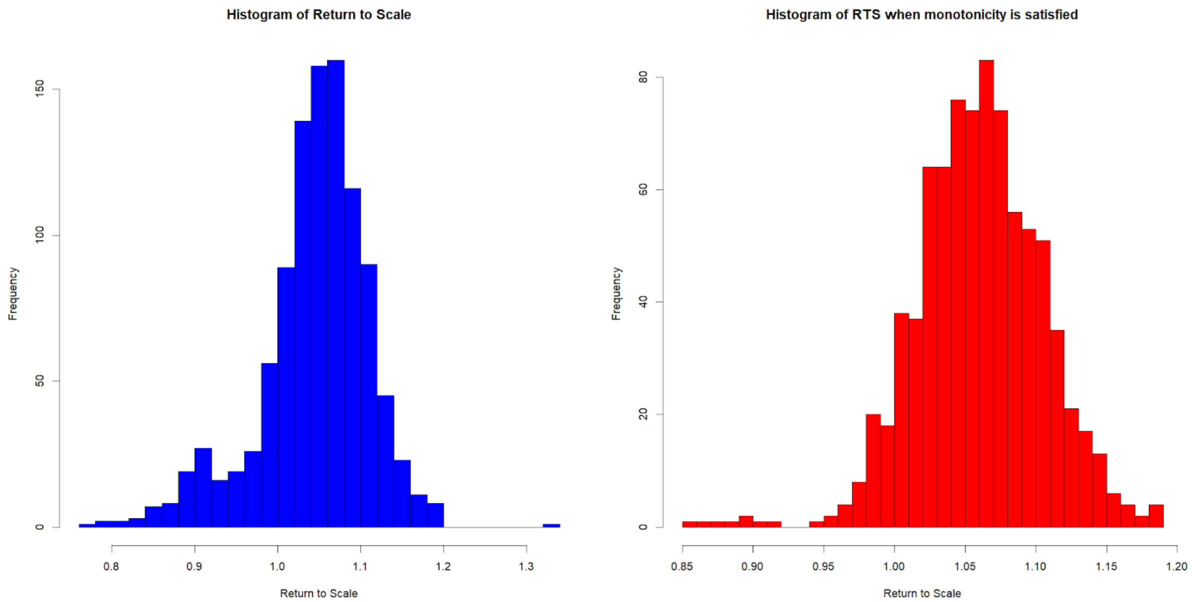
Figure 5.3



3.3.3 Returns to scale translog model

In figure 5.4 we have generated two histograms of return to scale. The blue side represents when all rice farms are included, and the red side includes only farms that satisfy the monotonicity condition. In the blue side the return to scale ranges from 0.8 to 1.2, with some deviation from the mean at 1.3. The red side ranges from 0.95 to 1.2 with some farms having a decreasing return to scale at 0.85 to 0.90. As observed on the blue side most farms lay around the 1.05 on the scale, which indicates to us that there is a slight increasing return to scale, but nothing major. The red side however the farms lay around 1.05 to 1.1, which suggest that there is an increasing return to scale, although small. The farms that satisfy the monotonicity condition tend to perform better, which is natural considering when the monotonicity condition is fulfilled, they use inputs like seed, phosphate, size, labor and urea efficiently and ensure that inputs contribute positively to output.

Figure 5.4



3.4 Goodness of fit analysis

Linear model:

```
[1] 0.8689767
```

Cobb Douglas model:

```
          [,1]
[1,] 0.836946
```

Quadratic model:

```
[1] 0.8935379
```

Translog model:

```
          [,1]
[1,] 0.8652055
```

Based on the R-squared value, the quadratic model shows the best fit and cobb douglas show the worst fit. It could be useful to look at the R-squared for each model using a logarithmic form because both the cobb douglas and the translog model uses it.

Linear model:

```
          [,1]
[1,] 0.8001924
```

Cobb Douglas model:

```
[1] 0.8761875
```

Quadratic model:

```
          [,1]
[1,] 0.8466952
```

Translog model:

```
[1] 0.8877048
```

When we do the linear and quadratic models with logarithms, we can see that the translog model is the best fit and that the linear model is the worst fit.

RESET test

Linear:

```
	RESET test

data:  linear_model
RESET = 20.545, df1 = 2, df2 = 1017, p-value = 1.79e-09
```

Cobb Douglas:

```
	RESET test

data:  cobb_douglas
RESET = 2.0134, df1 = 2, df2 = 1018, p-value = 0.1341
```

Quadratic:

```
	RESET test

data:  quadratic_model
RESET = 8.3517, df1 = 2, df2 = 1002, p-value = 0.0002528
```

Translog:

```
	RESET test

data:  translog_model
RESET = 0.9575, df1 = 2, df2 = 1003, p-value = 0.3842
```

Based on the results from the reset test for each model, we see that the linear and quadratic models are rejected because of the low p-value. The cobb douglas and translog model however can be kept as there is no statistically significant p-value to reject them.

4. Cost function

4.1 Short run cost function

In this chapter, we are going to estimate a short-run cost function for rice production. Since we do not have price data for land (size), we will treat it as a fixed input, justifying a short-run analysis more appropriate for our dataset.

A short run cost function represents the relationship between total production cost and output, assuming that at least one input is fixed, as in our case this will be size. Therefore, only variable inputs can be adjusted in the short run to influence the production level of the rice farms. Our fixed input is therefore land, while our variable costs would be seed, urea, phosphate and labor. The reason why we use these variables is they can fluctuate significantly in the short term, depending on production needs. This analysis can evaluate the cost efficiency for the rice farms, and help making better short-term production decisions (Kumar, 2023)

Size is our quasi-fixed input, and labor, seed, urea and phosphate is our total variable cost. We add our variables together and multiply them against the price in R to get our total variable cost.

```
# Calculate Total Variable Cost (TVC)
logmodel$TVC <- logmodel$wage * logmodel$totlabor +
  logmodel$pseed * logmodel$seed +
  logmodel$purea * logmodel$urea +
  logmodel$pphosp * logmodel$phosphate
```

We make our inputs in log format to estimate the following cobb-douglas short run cost function, based on our data.

```
Call:
lm(formula = log_TVC ~ log_wseed + log_wurea + log_wlabor + log_wphosphate +
    log_size + log_output, data = logmodel)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01284 -0.20998  0.00447  0.20932  1.20843

Coefficients:
               Estimate Std. Error t value             Pr(>|t|)
(Intercept)     3.35167    0.51226   6.543      0.0000000000953 ***
log_wseed       0.02093    0.03261   0.642               0.5211
log_wurea      -0.41407    0.37401  -1.107               0.2685
log_wlabor      0.77028    0.04042  19.056 < 0.0000000000000002 ***
log_wphosphate  0.75417    0.36589   2.061               0.0395 *
log_size        0.45232    0.02659  17.011 < 0.0000000000000002 ***
log_output      0.38056    0.02563  14.848 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3136 on 1019 degrees of freedom
Multiple R-squared:  0.8987,    Adjusted R-squared:  0.8981
F-statistic:  1506 on 6 and 1019 DF,  p-value: < 0.00000000000000022
```

Price for seed, urea, labor and phosphate is the variable cost. We see that only the price for labor and phosphate is statistically significant at a 5% level. Labor and phosphate both have respectively high coefficients at 0.77 and 0.75. This means that a 1% increase in labor price or price of phosphate would increase the total variable cost at 0.77% and for phosphate 0.75%. Urea has a negative coefficient at -0.41 though not significant at the 5% level, which may occur since urea has high multicollinearity in our data. This suggests that higher urea prices are associated with lower total variable costs, which is counterintuitive. It may also be that urea is already used in excess amount, therefore marginal productivity may decrease. The relative cost in the short-run cost function could in theory lead to not getting a higher rice output. Urea also has a variance inflation factor of 16.9, which is quite high, so the problem is most likely coming from high multicollinearity. Seed has a high p-value as well at 0.5211. If we sum up our coefficients of our variable input in prices we get (0.02093+0.77028+0.75417-0.41407= 1.13131), making our short term-term cost function not linearly homogeneous, as they do not equal to 1. This means that if we were to double our input prices, the total variable cost for rice farms would increase by more than double, 1.131311 instead of 1. If we were to interpret this it means that rice farms have increasing return to scale in prices, meaning if all input prices rise proportionality, total variable cost would increase more than proportionality. Output has a coefficient of 0.38 and a p-value of <0.05. This means a 1% increase in output leads to a 0.38% increase in total variable cost, which makes sense to think that when output increases, variable cost must also increase.

Size however shows a coefficient of 0.45, which contradicts microeconomic theory. Just like Henningsen (2024) experienced, we see that our fixed input size also has a coefficient that is both positive (0.45) and significant, suggesting that increasing land size leads to an increase in variable cost. The monotonicity principle expects fixed inputs not to increase the variable costs when output is held constant, therefore violating the condition.

4.2 Applying shephard's lemma

We are going to calculate the partial derivative of the short term cost function with respect to price of each input, that will give us the demand function for each variable input.

Our short-run cost function looks like this:

$$Log(TVC) = a + b_1 \log(w_{labor}) + b_2 \log(w_{urea}) + b_3 \log(w_{seed}) + b_4 \log(w_{phosphate}) + b_5 \log(size) + b_6 \log(output)$$

Where TVC is the total variable cost. Labor, urea, seed and phosphate are the prices, size is our fixed input and log(output) represents the rice production.

```
  demand_labor      demand_urea         demand_seed          demand_phosphate
 Min.   :1.264   Min.   :-1.2707   Min.   :0.03259   Min.   :1.311
 1st Qu.:1.714   1st Qu.:-1.0233   1st Qu.:0.04326   1st Qu.:1.645
 Median :1.847   Median :-0.9636   Median :0.04664   Median :1.752
 Mean   :1.857   Mean   :-0.9671   Mean   :0.04677   Mean   :1.759
 3rd Qu.:1.981   3rd Qu.:-0.9052   3rd Qu.:0.04976   3rd Qu.:1.860
 Max.   :2.649   Max.   :-0.7198   Max.   :0.06502   Max.   :2.314
>
```

The average demand for labor is 1.86, seed is 0.047 and phosphate is 1.76. Note that the demand for urea is negative, at a mean value of -0.967. This might originate from the fact that our model is not linearly homogeneous; the monotonicity condition is not fulfilled in a cobb-douglas specification as we see in the short-run cost function. It might also indicate that some farms may overuse urea without it giving any noticeable yield in production, therefore our model may perceive urea as a less important input although not significant. Our dataset also has 4 different seasons of rice production, this also can contribute to the uncertainty for the rice farms. When farms experience dry periods, it might result in overusing urea in hopes to maintain the same level of production as in more favorable periods.

4.3 Cost minimizing behaviour

As we have seen above in our cost minimizing behaviour for our cobb-douglas function in chapter 2, we study the MRTS between different inputs against the price ratio. We see that the cost minimizing behaviour deviates for most of the farms away from the effective line between MRTS and price ratio. This means that most farms have an ineffective use of resources, therefore most farms could use less input to maintain the same level of rice production. Most farms use urea suboptimal, therefore it supports our finds in the short run cost function where we have a negative coefficient. Labor seems to be mostly overused against urea and phosphate, however underused against seed.

5 Discussion

In this paper we have analysed the rice production in Indonesian farms by using three different production functions to determine the best fit for predicting rice yield in the rice farms dataset.

Through our analysis, the Cobb Douglas and translog model stood out as good fits with high R-squared values to explain the variance in output. The translog model however performed the best when doing the wald test and likelihood ratio test.

The short run cost function revealed that only the labor and phosphate had a statistically significant impact on the variable costs and had high elasticities. Urea’s effect happened to be negative, which could potentially mean overuse or being used inefficiently. Urea may be used inefficiently due to that some farms may increase urea, therefore reduce the need to other inputs. A factor to consider also is that the marginal productivity for urea as seen in figure 5.3 is quite low compared to the other inputs, almost having diminishing returns, it could lead to a less important role in production as it increases. We suspect that though other arguments are valid, the main reason is due to multicollinearity in the data.

Limitations to the paper. The dataset does not contain cost price over land, which limited our ability to make a long run cost function. The dataset had seasonal variations, which could impact the efficiencies of certain inputs, for example in dry periods where urea may have been overused. Maybe including weather data could be useful for being able to analyze this dataset better.

Based on our findings, we recommend the translog model for further analysis of this dataset, as it is flexible and accounts for input interactions. For the farm managements, there is shown to be a clear benefit in optimizing their input levels, especially by addressing the observed inefficiency in labor and urea. More efficient input strategies could improve the productivity and profits for the Indonesian rice farms.

Conclusion

We have highlighted the possibility for Indonesian rice farms to improve productivity and reduce their costs by optimizing their input factors, most importantly labor and phosphate. The translog model seems the most robust of the different production functions.

References:

Erwidodo. (1990). *Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in West Java, Indonesia* (Doctoral dissertation, Michigan State University. Department of Agricultural Economics).

Store norske leksikon (2005-2007): *urea* i *Store norske leksikon* på snl.no. Hentet 21. november 2024 fra https://snl.no/urea

Feng Q, Horrace WC (2012). "Alternative technical efficiency measures: Skew, bias and scale." *Journal of Applied Econometrics*, **27**(2), 253-268. doi:10.1002/jae.1190.*

Kumar, Prince(2023). Short run cost function. *Managerial Economics.* *https://themba.institute/managerial-economics/short-run-cost-function/*

Henningsen, A. *(2014).* Introduction to econometric production analysis with R. Department of Food and Resource Economics, *University of Copenhagen.* *https://zenodo.org/records/11093657*

Tapas Kundu (2024). Linear Model in Economics, *Norges Arktiske Universitet:*

https://uit-sok-3011-h24.github.io/tapaslectures/lecture4/apa4.pdf

Dataset used: Indonesia ricefarms: https://vincentarelbundock.github.io/Rdatasets/doc/plm/RiceFarms.html

The use of AI in this assignment

We used the AI ChatGPT by openAI.

Level of use: AI-assisted Idea generation and structuring:

We used it primarily to help us to be more effective with the R code, and were used to provide some interpretation regarding R code and explanation for our own understanding. All final analysis and results in the assignment reflect our own understanding and interpretation, with ChatGPT used primarily as a supporting tool for R-code and interpretation

8. Appendix

Appendix is in the QMD file, where we show plots and explains the functions.