CHAPTER 6

ANALYSIS & FINDINGS: STATISTICAL PRIORITISATION OF OIL PALM VARIABLES

6.1 Introduction

This chapter looks at the prioritisation of variables in the palm oil industry from the environmental, social and economic sectors. Most previous studies have focused on one or two sectors but this study covers variables across all 3 sectors to develop a holistic and more multi-faceted assessment of palm oil's sustainability.

The study analysed the relationship between total planted area and palm oil price using the 120 variables previously selected in Chapter 5. Total planted area and palm oil price were selected as an outcome of the perception survey in Chapter 4 to be most important determinants for the oil palm industry. Data for the total planted area (hectares) for oil palm from 1975 to 2010 and the average annual price of CPO (local delivery) from 1980 to 2010 (RM/Tonne) were used. The time series analysis that follows will help to show which values are important for Malaysia as determinants of palm oil's sustainability.

6.2 Limitations of the Analysis

As with all economic tests, the following limitations were observed for this study:

1. Only micro data were selected as important oil palm variables for Malaysia. Macro country data based information such as GDP and FDI were not considered;

2. Two dependent variables were selected from the MPOC database: palm oil price per year and total planted area for oil palm in Malaysia. These dependent variables both link to two important determinants identified during the perception survey of the stakeholders;
3. Pearson and Spearman correlation tests were carried out to select significant variables, and the Pearson test was found to be more suitable. Based on the correlation test, the list of significant variables was established;
4. The findings from the Pearson correlation test were then used for the Linear Regression test. Both the Enter and Stepwise methods were used for the Regression exercise, and it was found that Stepwise eliminated all the variables due to multi-co-linearity. And therefore, the Enter methods found to be more suitable, and variables with least proxies were taken to present important measurement;
5. Environmental and plantation variables were found to highly correlate or "redundant" with one another, and exchange capacity are very tightly correlated since, any one of these variables could be used as a proxy for all the others. Generally, it was best to choose the variables which were most likely to be the direct cause of a response, and/or a variable which has been used in other ecological studies.

(Note: normality test, multi-co-linearity, homoscedasticity and independence of error terms were all observed.)

6.3 The List of Variables Chosen and Tested

Not all the variables identified could be used to run the regression and correlation test as the data records/fields requirements were not met, as the data sets were not complete for statistical analysis. These variables were tested for significance against the palm oil

price and total planted area in Malaysia, to show if they have any impact and could change the production patterns. The final selection of variables for testing is listed in ***Table 6.1.***

Table 6.1: List of Variables Selected for Testing

Variables
Agricultural land (% of land area)
Agricultural methane emissions (% of total)
Agricultural nitrous oxide emissions (% of total)
Agriculture, value added (% of GDP)
Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)
Annual freshwater withdrawals, total (billion cubic meters)
Arable land (% of land area)
Arable land (hectares per person)
CO2 emissions (kt)
CO2 emissions (metric tons per capita)
Employment in agriculture (% of total employment)
Fertiliser consumption (kilograms per hectare of arable land)
Forest area (% of land area)
Forest area (sq. km)
Methane emissions (kt of CO2 equivalent)
Nitrous oxide emissions (thousand metric tons of CO2 equivalent)
Organic water pollutant (BOD) emissions (kg per day)
Permanent cropland (% of land area)
Terrestrial protected areas (% of total surface area)
Terrestrial protected areas (number)
Area Under Oil Palm (Mature & Immature) - P.Malaysia
Area Under Oil Palm (Mature & Immature) – Sabah
Area Under Oil Palm (Mature & Immature) – Sarawak
Principal Statistics of Oil Palm Estates (number of estates)
Principal Statistics of Oil Palm Estates (planted hectares)
Principal Statistics of Oil Palm Estates (production of fresh fruit bunches)
Principal Statistics of Oil Palm Estates (yield per hectare)
Principal Statistics of Oil Palm Estates (local delivered average price)
Principal Statistics of Oil Palm Estates (total number of workers employed)
Total Planted Hectares of Oil Palm (Total Area)
Total Planted Hectares of Oil Palm (Estate)
Total Planted Hectares of Oil Palm (Smallholding)
Annual Average Prices of Oil Palm Products(CPO - local delivered)
Annual Average Prices of Oil Palm Products (PK - Ex-Mill)
Annual Average Prices of Oil Palm Products (FFB - 1% Extraction Rate)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Peninsular Malaysia)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sabah)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sarawak)

Source: Author, 2013

6.4 Analysis and Findings – Significant Variables for the Industry

All the variables selected were tested for their significance against the two dependent variables: palm oil price and the total planted area in Malaysia. The correlation results are shown in Table 6.2. Details of the correlation exercise for the variables are found in ***Appendix 3***.

Table 6.2 has been divided into variables that were found to be significant palm oil price and total planted area through the bi-variant test without conversion of the data into normal logarithm.

Table 6.2: List of Significant Variables for Palm Oil Price and Total Planted Area

Palm Oil Price	Total Planted Area
Principal Statistics of Oil Palm Estates (harvested hectares)	Agricultural land (% of land area)
Principal Statistics of Oil Palm Estates (production of fresh fruit bunches)	Arable land (% of land area)
Principal Statistics of Oil Palm Estates (total number of workers employed)	Employment in agriculture (% of total employment)
Principal Statistics of Oil Palm Estates (number of estates)	Fertilizer consumption (kilograms per hectare of arable land)
Principal Statistics of Oil Palm Estates (yield per hectare)	Forest area (sq. km)
Principal Statistics of Oil Palm Estates (planted hectares)	Terrestrial protected areas (% of total surface area)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Peninsular Malaysia)	Area Under Oil Palm (Mature & Immature) –Peninsular Malaysia
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sabah)	Area Under Oil Palm (Mature & Immature) –Sabah
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sarawak)	Area Under Oil Palm (Mature & Immature) –Sarawak
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Malaysia)	Area Under Oil Palm (Mature & Immature) –Malaysia
Total Planted Hectares of Oil Palm (Total Area)	Principal Statistics of Oil Palm Estates (number of estates)

……Continue

Table 6.2: List of Significant Variables for Palm Oil Price and Total Planted Area

Palm Oil Price	Total Planted Area
Agricultural land (% of land area)	Principal Statistics of Oil Palm Estates (planted hectares)
Arable land (% of land area)	Principal Statistics of Oil Palm Estates (harvested hectares)
Employment in agriculture (% of total employment)	Principal Statistics of Oil Palm Estates (production of fresh fruit bunches)
Fertilizer consumption (kilograms per hectare of arable land)	Principal Statistics of Oil Palm Estates (yield per hectare)
Forest area (sq. km)	Principal Statistics of Oil Palm Estates (local delivered average price)
Terrestrial protected areas (% of total surface area)	Principal Statistics of Oil Palm Estates (total number of workers employed)
Area Under Oil Palm (Mature & Immature) –Peninsular Malaysia	Annual Average Prices of Oil Palm Products (CPO - local delivered)
Area Under Oil Palm (Mature & Immature) –Sabah	Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Peninsular Malaysia)
Area Under Oil Palm (Mature & Immature) –Sarawak	Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sabah)
Area Under Oil Palm (Mature & Immature) –Malaysia	Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sarawak)
	Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Malaysia)

Source: Author, 2013

When the data was converted to natural logarithm form, for each of the variables and tested for significance, the list of significant variables reduced in number and the once found to be as in ***Table 6.3*** for palm oil price and ***Table 6.4*** for total planted area. Normalized data showed considerably fewer variables as important, as seen in ***Tables 6.3*** and ***6.4.*** Thereby, when the data is normalized, the number of significant variables becomes fewer.

Table 6.3: Significant Variables Converted into Natural Logarithms for CPO local Delivery.

Significant to Average Annual Prices of Oil Palm Products
Principal Statistics of Oil Palm Estates (harvested hectares)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sabah)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sarawak)
Agricultural land (% of land area)
Arable land (% of land area)
Employment in agriculture (% of total employment)
Fertiliser consumption (kilograms per hectare of arable land)
Area Under Oil Palm (Mature & Immature) – Peninsular Malaysia
Area Under Oil Palm (Mature & Immature) – Sabah
Principal Statistics of Oil Palm Estates (number of estates)
Principal Statistics of Oil Palm Estates (planted hectares)

Source: Author 2013

The correlation test for this study was vital to reduce the number of data that were significant, so that the next level of test (the regression exercise). This was necessary to reduce the impact of multi-co-linearity that came with the industry data being very closely link to each other through its natural state. However, discretion of the researcher and other statisticians were used to run the linear equations for the next level of test.

Table 6.4: Significant Variables Converted into Natural Logarithms for Total Planted Area.

Significant to Total Planted Hectares of Oil Palm (Total Area)
Agricultural land (% of land area)
Arable land (% of land area)
Principal Statistics of Oil Palm Estates (number of estates)
Principal Statistics of Oil Palm Estates (planted hectares)

Source: Author, 2013

6.5 Previous Studies that Support a Regression Exercise

The regression work was based on the general long-run model, written as following:

$$production = f(area, price) \quad (1)$$

Where production represents the production of palm oil in Malaysia; area represents the total area planted, price represents the price of palm oil within a given period time trend *t*,. The subscript, 't' in each variable represents time lags, trend.

Studies previously conducted by Asari, et al. (2011) and Njoo, 2001 were considered. The study by Asari, et. al, 2011 showed the implications of land area and price on the production of palm oil where three variables are then transformed into the constant variance. This was expressed in linear logarithm form; i.e. ln production (log of production of Malaysian palm oil), ln area (log of total area planted) and ln price (log of palm oil price). The multivariate relationship represented by equation (2) was investigated by Asari et al. in their paper:

$$\ln(production_t) = \beta_0 + \beta_1 \ln(area_t) + \beta_2 \ln(price_t) + \varepsilon_t \quad (2)$$

A complex, multi-faceted work study carried out by Jamal Othman (2003) linked agricultural trade, land demand and environmental externalities for the oil palm industry in South East Asia. The analysis of agricultural trade-environmental linkages requires knowledge of: i) the underlying agricultural production function which links factor use, type of technology and output supply, ii) how the demand for output is affected by shifts in export demand and related trade policies imposed by producing and competing

countries, and iii) how changes in production factors, particularly land, result in externalities, in this case the production of haze. The following equations illustrate these linkages: Demand for agricultural output by equation (3) Equation (3) expresses the demand for agricultural output (q_O^D) which is a function of domestic demand $(p_O^D(d_O^D)$ and export demand $(p_O^E(d_O^E))$:

$$q_O^D = f(p_O^D(d_O^D) + p_O^E(d_O^E)) \quad (3)$$

This study employed single and multi-country output supply exogenous policy models with explicit factor markets to examine agricultural land demand-trade linkages in the world vegetable oil markets (Jamal, 2003).

6.6 Limitations for Regression Test and Justification

To establish important variables that impact or have implications on production, the limitations listed below were used when carrying out the regression tests:

a) Dependent Variable: Palm Oil Price and Total Planted Area
b) Independent Variable: As per Table
c) Test run: Linear Regression (Enter and Stepwise), SSPS software
d) Data Source: Malaysian Palm Oil Council, Malaysian Palm Oil Board, Department of Statistics, Malaysia, World Bank Databank, International Monetary Fund Data, etc.
e) Time Frame: 1980-2010

f) Plantation data were more natural in their form, mostly lacking in monthly figures or information.

6.7 Regression Exercise and Limitations

As the intention of the study was only to seek to prioritise the list of important variables for sustainability development, the multivariate relationship represented by equation 2 was investigated by the researcher with the dependent variables being total planted area and palm oil price, per equation 4 and 5:

$$\ln(totalplantedarea_t) = \beta_0 + \beta_1 \ln(area_t) + \beta_2 \ln(price_t) + \varepsilon_t \quad \textbf{(4)}$$

$$\ln(palmoilprice_t) = \beta_0 + \beta_1 \ln(area_t) + \beta_2 \ln(price_t) + \varepsilon_t \quad \textbf{(5)}$$

As the earlier correlation test produced two sets of significant independent variables, and the regression exercises were carried out for palm oil price and total planted area, these are depicted in the ***Figure 6.1*** and ***Figure 6.2.*** The figures are designed to show the need for variables to be priorities for both palm oil price and total planted, and list important variables for the industry.

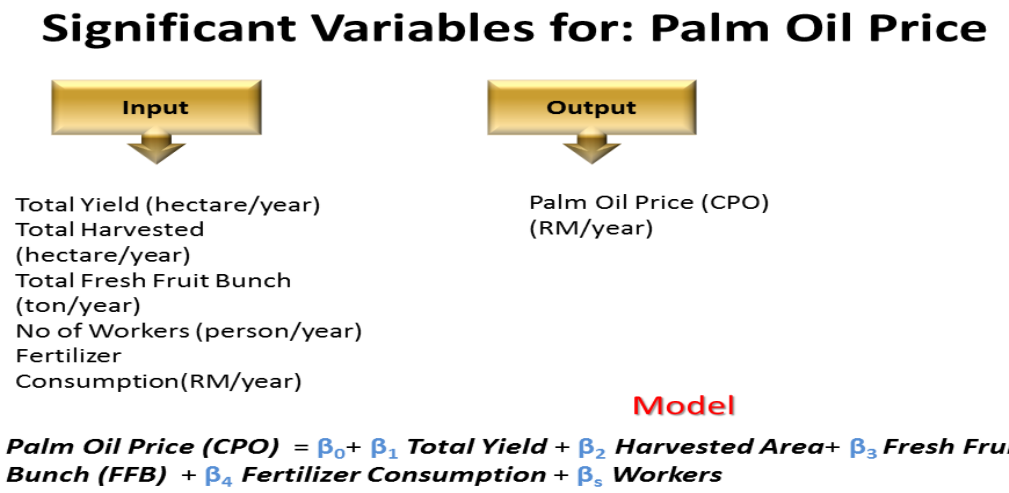

Figure 6.1: Variables tested for Palm Oil Price

Initial test for palm oil price were carried out with the following variables: total yield, total harvested area, fresh fruit bunch harvested, fertilizer consumption and number of workers in the industry.

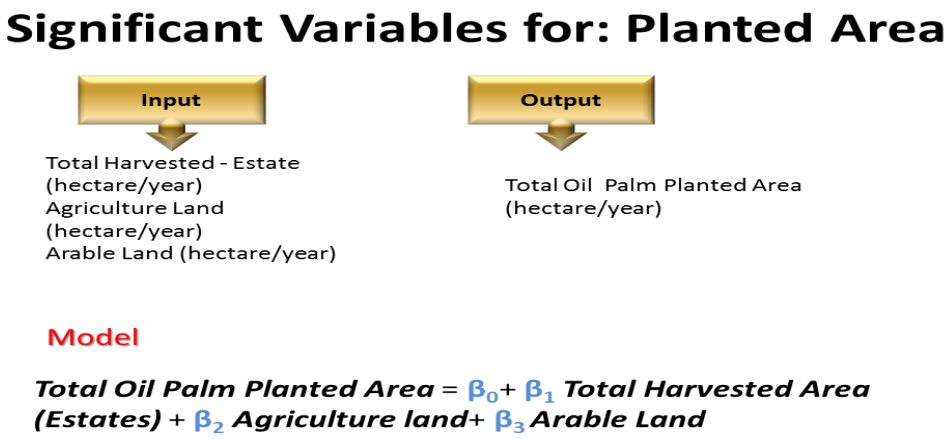

Figure 6.2: Variables tested for Total Planted Area

Initial test for total planted area were carried out with the following variables: total harvested area, total agriculture land available and total arable land available.

6.8 Results from the Regression Exercise for Total Planted Area

For total planted area, the model summary is described in ***Appendix 3***, and the regression values are in ***Table 6.5***.

Table 6.5: Results from Regression for Total Planted Area

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.288	.136		2.121	.041
	lnagricultural	.573	.124	.118	4.641	.000
	lnarable	-1.423	.291	-.195	-4.892	.000
	lnemployment	.016	.017	.008	.911	.368
	lnfertilizer	.025	.014	.021	1.739	.090
	lnforestarea	-.013	.009	-.026	-1.504	.141
	lnterrestrial	.081	.062	.039	1.311	.198
	lnareapm	.016	.083	.036	.197	.845
	lnareasabah	.179	.181	.354	.987	.330
	lnareasrwk	-.256	.152	-.462	-1.681	.101
	lnareamsia	.033	.064	.073	.512	.612
	lnnumberestate	.330	.055	.204	5.987	.000
	lnplanted	.837	.031	.828	27.138	.000

a. Dependent Variable: lntotalarea

Source: Author, 2013

The regression test shows that there are four variables out of the 17 of the independent variables which are significant at 0.05 (Sig. level) and these four independent variables test significant for the Total Planted Hectareage of Oil Palm, and they are:

I. Agricultural land (% of land area),

II. Arable land (% of land area),

III. Principal Statistics of Oil Palm Estates (number of estates),

IV. Principal Statistics of Oil Palm Estates (planted hectares)

6.9 Results from the Regression Exercise for Palm Oil Price

For palm oil price, the model summary is described **in Appendix 3**, and the regression values are in ***Table 6.6*** (The correlation variables are significant after transformation.)

Table 6.6: Results from Regression for Palm Oil Price

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.324	1.120		7.434	.000
	LNHARVES	16.484	2.450	13.923	6.729	.000
	LNPRODUC	-.132	.108	-.287	-1.226	.227
	LNYIELD	7.040E-02	.523	.027	.135	.894
	LNLABOR	-.279	.104	-.392	-2.674	.011
	LNFFBPM	7.716	4.245	3.179	1.818	.076
	LNFFBSAB	-4.896	3.300	-2.079	-1.483	.146
	LNFFBSRW	-3.087	2.434	-1.178	-1.268	.212
	LNNUMBER	5.008	.877	2.636	5.710	.000
	LNPLANTE	-18.727	2.485	-15.774	-7.536	.000

a. Dependent Variable: LNCPO

Source: Author, 2013

Principal Statistics of Oil Palm Estates (harvested hectares), Employment in agriculture (% of total employment) and Principal Statistics of Oil Palm Planted Area (Number of Estates) and Principal Statistics of Oil Palm Estates (planted hectares) are significant ($p<0.05$) and show that they affected the Average Annual Prices of Oil Palm Product (CPO-local delivered).

6.10 Proxies for the Representative Variables

The main limitation t of this work was found to be that the environmental and plantation variables are highly correlated or "redundant" with one another, and exchange capacities are very tightly correlated. Thus, any one of these variables could be used as a proxy for the others. (The regression details are in ***Appendix 2***).

The following variables are proxies for total planted area and total harvested area in the regression work for Palm Oil Price: ***Table 6.7***

Table 6.7: Proxies Variables for Palm Oil Price Regression Test

Proxies for Palm Oil Price
Principal Statistics of Oil Palm Estates (harvested hectares)
Principal Statistics of Oil Palm Estates (production of fresh fruit bunches)
Fresh Fruit Bunches (FFB) Yield Tonnes/Hectare (Sabah)
Total Planted Hectares of Oil Palm (Total Area)
Area Under Oil Palm (Mature & Immature) – Peninsular Malaysia
Area Under Oil Palm (Mature & Immature) – Sabah
Area Under Oil Palm (Mature & Immature) - Sarawak

Source: Author, 2013

The following variables are proxies for planted area and harvested area in the regression work for Total Area Planted: ***Table 6.8***

Table 6.8: Proxies Variables for Total Planted Area Regression Test

Proxies for Total Planted Area of Oil Palm in Malaysia
Area Under Oil Palm (Mature & Immature) – Peninsular Malaysia
Area Under Oil Palm (Mature & Immature) – Sabah
Area Under Oil Palm (Mature & Immature) – Sarawak
Area Under Oil Palm (Mature & Immature) – Malaysia
Principal Statistics of Oil Palm Estates (number of estates)
Principal Statistics of Oil Palm Estates (planted hectares)
Principal Statistics of Oil Palm Estates (harvested hectares)

Source: Author, 2013

6.11 The Final Model Set-up and Selection of the Most Important Variables

The Principal Statistics of Oil Palm Estates (harvested hectares) and Principal Statistics of Oil Palm Estates (local delivered average price) are significant ($p>0.05$). They show that harvested hectares and local delivered average price in Malaysia affect the Average Annual Prices of Oil Palm Product (CPO-local delivered). The first model set-up for palm oil price can be depicted as:

$$\textit{Palm Oil Price (CPO)} = \beta_0 + \beta_1 \textit{ Harvested Area} + \beta_3 + \beta_2 \textit{ Local Delivery Price}$$

Four independent variables significantly affect the Total Planted Hectareage of Oil Palm (Total Area): Agricultural land (% of land area), Arable land (% of land area), Principal Statistics of Oil Palm Estates (number of estates), and Principal Statistics of Oil Palm Estates (planted hectares). The second model set-up for total planted area of palm oil in Malaysia can be depicted as:

$$\textit{Total Oil Palm Planted Area} = \beta_0 + \beta_1 \textit{ Planted Hectares} + \beta_2 \textit{ Number of Estate} + \beta_4 \textit{ Agricultural Land} + \beta_s \textit{ Arable land}$$

From both these models, it can now be confirmed that for oil palm plantations in Malaysia, the important determinates for sustainable management, using the above correlation matrix and the two models established, the most important variables for the management of oil palm plantations in Malaysia are:

I. *Planted Hectares or Total Planted Area,*
II. *Harvested Area,*
III. *Agricultural Land Area,*

IV. *Arable Land Area, and*

V. *Local CPO Delivery Price.*

For each of these variables, the concept may vary between Malaysia and other territories. In Malaysia, the following contexts apply:

I. Planted Hectares or Total Planted Area: land cover in hectares for oil palm crop, in mature and immature state for Peninsular Malaysia, Sabah and Sarawak.

II. Harvested Area: land cover with mature oil palm crop that can be harvested in Peninsular Malaysia, Sabah and Sarawak.

III. Agricultural Land Area: land in Malaysia that is available for agricultural use, which includes the planting of oil palm and other crops e.g. rubber, tea and paddy.

IV. Arable Land Area: land that can be used for growing crops (being fertile and of a certain soil type). Due to development, some of this land use may have been converted to housing, roads or other land use.

V. Local CPO Delivery Price: price of crude palm oil that is delivered locally from the mills in Malaysia. (This price may vary according to geographic regions in Malaysia.)

6.12 Discussion on the Findings: Prioritization of Variables

The analysis in this chapter represents the domestic structure of the important variables for oil palm plantations in Malaysia. In this context, the following findings are highlighted:

a. Difficulty in data procurement: even though the oil palm industry is the largest agricultural sector in Malaysia and has been highly institutionalized, for

sustainable development the availability of data for research and development is the greatest hurdle faced.

b. Lack of important sustainability information: Some variables could not be tested, even though they have been deemed important. The data for these variables were not sufficient quantitatively for economic testing to be carried out. Furthermore, the information has become newly important and the data for values such as for greenhouse gas emissions, carbon footprints, etc are not available for earlier periods.

c. The significant results for total planted area and oil palm price were established via the perception survey carried out, and were further confirmed through correlation and regression analysis. Both these variables are the main links between the perception and economic analyses.

d. Even though the overall model that was set up for this analysis was simple, it was crucial to display the prioritization of variables that were necessary for the consideration of sustainability measurements in Malaysia, for the development of the palm oil industry.

e. In terms of the regression findings for total planted area, it can now be confirmed that the important variables are Agricultural land (% of land area), Arable land (% of land area), Principal Statistics of Oil Palm Estates (number of estates), and Principal Statistics of Oil Palm Estates (planted hectares).

f. For the regression findings for oil palm price, it can now be confirmed that the important variables are Principal Statistics of Oil Palm Estates (harvested hectares) and Principal Statistics of Oil Palm Estates (local delivered average price).

6.13 Overall Statement

To develop the oil palm industry, both the available data and measurement of new variables should be given priority. Without the transparency of information between agencies in the country, this development will be hampered. Gaps and controversies cannot be overcome and there is no room for improvement.

Consistent efforts should also be made to measure missing data and set up proper databases. With both data transparency and data gaps managed well, gaps in sustainability measurement for Malaysia will not hinder further progress and development.