Statistical Analysis on Variable Cost at Aurubis

11/30/2014 JeongHoon Shin

Table of contents

Chapter 1. Introduction … 3

Chapter 2. Materials and Methods … 3

Chapter 3. Data Analysis … 5

Chapter 4. Interpretation and Conclusion … 6

Appendix … 7

Chapter 1. Introduction

Objective

Aurubis, the company that recycles and produces copper products wants to understand the business factors for effective cost management. Specifically, the company currently pays close attention to the relationship between the total variable cost and total production volume. According to its initial analysis it is not simply a deterministic function. Although in general variable cost is linearly related with production amount, the production volume in Aurubis' sales data did not sufficiently explain the variability of the variable cost and hence there needed to be an analysis on what might affect the relationship between the two. As an example, a simple linear regression model between the variable cost and the production volume showed that this blurred the relationship, showing a big standard error of the regression value and low R square value of 0.4, which means only approximately 40% of the variability of the variable cost is explained by the obvious factor, production volume.

One example easily presumed to be one of the causes for this discrepancy is 'economies of scale', meaning cost advantage that arises with increased output of a product because of the inverse relationship between the quantity produced and per-unit fixed costs. Often the greater the quantity of a good produced, the lower the per-unit fixed cost because these costs are shared over a larger number of goods.

To find the factors like this and to obtain a better understanding of the factors and the variation of the variable cost are the project goal, ultimately for the company to make an optimal business plan.

Chapter 2. Materials and Methods

Considering the nature of the interested response and the aspects the company is interested in, regression modeling is chosen as the most appropriate method.

Before proceeding, part of the data is presented for readers' better understanding.

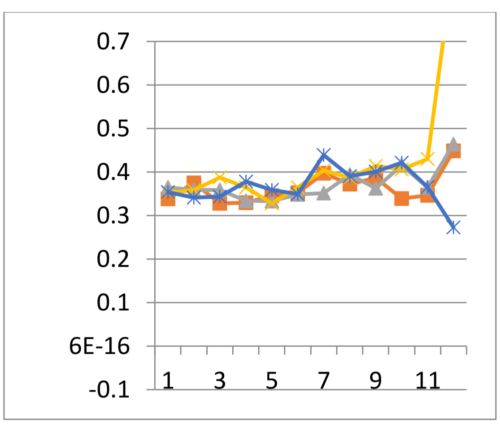
Year Month	(sold quantity) Cust Lbs	(sold quantity)Frank Lbs	Shutdown	total cost varialbe ($)	total fixed cost / lb ($)	average variable cost / LB ($)	Time	Cust_%	production Total Lbs	Shutdown	production Total Lbs (millions)	Copper, grade A cathode Monthly Price	Copper LME Settlement Price
Jan-12	10000	8000	0	4000	1200	0.222	0	0.556	18,000	0	18	3,168.10	
Feb-12													
Mar-12													
Apr-12													
May-12													
Jun-12													
Jul-12													
Aug-12													
Sep-12													
Oct-12													
Nov-12													
Dec-12													
Jan-13													
Feb-13													
Mar-13													
Apr-13													
May-13													
Jun-13													
Jul-13													
Aug-13													
Sep-13													
Oct-13													

THE VALUES IN THE FIRST ROW ARE FICTIONAL.
OTHER SPECIFIC VALUES ARE DELETED FOR SECURITY
AS REQUESTED BY THE COMPANY

Sheet5 | VC | AVC model | AVC model from SPSS | AVC model with more variables | Monthly AVC pattern | Correl ...

Figure 1 Partial data

The data set was ranged from the date of Jan 2005 to May 2014. However, as there had been a clear gap in the variable cost trend between years as following graphs show, data from recent years (from 2012) were mainly analyzed to reflect the recent trend.



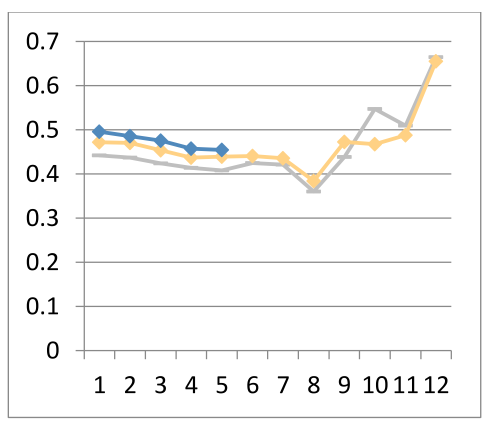


Figure 2 Variable cost from 2005 ~ 2009

Figure 3 Variable cost from 2012 ~ 2014

Here are a few explanations of the terms used in this report.

(1) Frank LBS, Cust LBS : Represents the amount of shipping pounds to Franklin, the biggest client, and the amount of shipping pounds to all the other clients.
(2) Average Variable Cost (AVC) : (Cust Lbs + Frank Lbs) / Total variable cost
(3) Shutdown : A situation when the copper price soars and Aurubis cuts down on the amount of production. This lasts for one month, and often happens in September in each year. In the data, the value 1 in the column "Shutdown" indicates that shutdown happens.
(4) Total variable cost sometimes may be referred to as "TVC" and total production volume as "total volume" or "plant volume".

For this project, the questions to answer are defined as follows.

1. About the relationship : How does total production volume affect the total variable cost (and average variable cost and marginal cost)? Is it linear? How accurately can we estimate the effect of the relationships?
2. About synergy : Does the volatility of total production volume help explain more of the variation in total variable cost than just total production volume alone?

Chapter 3. Data Analysis[1]

Based on the company's interests, an initial model was formulated as follows.

[Total Variable Cost = a(total production volume) + b(volatility of production volume) + c(% Franklin) + d(time) ...,] where c and d were meant to control for factors such as product mix and inflation and a and b were the parameters to pay attention.

Thereafter, the research was progressed from this initial regression model with following specific steps.

1) Checking the assumptions of the multiple regression model.

   That is,

   - If the error terms are normally distributed with a mean of 0 at each predictor variable.

[1] The detailed steps are attached in Appendix.

- If the error terms have equal variances at each predictor variable.
- If the errors are independent.

2) Hypothesis test for F statistics to verify if the parameters are not zero, meaning the predictors have some actual relationship with the response variable, the Variable Cost.
3) Making the statistical models.
   Out of the methods for this, all possible regression modeling is adopted since there is not so many predictors.
4) Selecting the best ones that fit for the situation.
   For this some of the most relevant properties were taken into consideration including subject information, plots and some statistics.

As a result the final model created for this project was

Total cost variable = 1758.106 + 36.575*Time + 0.228*Production_total + 4091.679*%chg_in_FC_to_Prod + 2147.17*Volatility

Regression model for variable cost

with the production volume's linear relationship with the variable cost together with its volatility. Also, the effect of inflation can be quantified for its future prediction on the variable cost.

Chapter 4. Interpretation and Conclusion

Total cost variable = 1758.106 + 36.575*Time + 0.228*Production_total + 4091.679*%chg_in_FC_to_Prod + 2147.17*Volatility

Regression model for variable cost

From the model, a few conclusions can be drawn.

We can see the linear relationship between the production volume and total variable cost. It shows increasing one unit of production causes 0.228 increase in the variable cost. Also, the more the production volume changes over months, the more the variable cost increases. And if the forecast on the clients' demands is inaccurate and thus there is a lot of gap with the production volume, it will cost much more variable cost. Additionally, as the Time variable reflects, the inflation effect on the variable cost at Aurubis is approximately 36.57 increase in the cost.

One of the points to think about for further research, is the fact that due to the lack of data some factors that may be important to the variable cost were not taken into consideration. For example, the data on how "economies of scale" affect TVC, and on the changeovers waste. These are the factors presumably related to the variable cost. Also, more to think about for further research would include normalizing the variables, and generating models using the Stepwise methods with all of its submethods – Forward, Backward, Stepwise – and finding the commons and narrowing down the best of the models using the subject knowledge for this project.

Appendix

Here are the detailed processes to generate the model for this project.

I) Variable selection

Basically, as the goal of this project was to research the relationship between the variable cost and the production volume, these two variables were included in the most regression models, and some other variables to be considered were included or excluded in the regression models.

The initial regression model included Time, Cust%, Total_LBS(Total production amount in that month), and Shutdown as independent variables to affect the dependent, average variable cost.

SUMMARY OUTPUT				
Regression Statistics				
Multiple R	0.789117698			
R Square	0.622706741			
Adjusted R Square	0.604523934			
Standard Error	0.052108366			
Observations	88			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	4	0.371961101	0.092990275	34.24700702
Residual	83	0.22536839	0.002715282	
Total	87	0.597329491		
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.138967973	0.101090156	1.374693425	0.17292656
Time	-0.000252603	0.000415485	-0.607972605	0.544865679
Cust_%	0.521342575	0.116865913	4.461032001	2.54105E-05
Total Lbs	-3.32709E-06	2.7179E-06	-1.224137379	0.22436383
Shutdown	0.058787963	0.023793804	2.470725741	0.015530513

Figure 4 Initial regression model obtained from Excel

AVC (Average variable cost) = 0.1389 - 0.0002*Time + 0.5213*Cust% - 0.000003*Total_LBS + 0.0587*Shutdown

Equation 1 Initial regression model

First there had to be a rearrangement of the independent variables in this initial regression model because every independent variables had p-values higher than the significance level of 0.95, which was used as a standard significance level for this project. Therefore, the parameters of the corresponding variables were likely 0, meaning that the variables had no effects on the dependent variable. Thus, the variables Time, Cust_%, and Total Lbs in the initial regression model appeared to be not significant in this regression model. However, this does not mean all of these variables were impractical to explain the variability of the dependent variable, as the significances may differ according to combinations of the independent variables in a regression model.

On top of this, because a pair of highly correlated variables can invalidate a regression model making individual parameters useless, correlation tests for each pair of candidate independent variables were performed prior to modeling with those variables.

The variable Shutdown was removed for the whole procedures in the project as this research was of a linear regression modeling, which does not contain a categorical variable. However, the variability of the variable cost could be reflected as the values of the production, variable cost, and all of the other variables represented the changes when the shutdown period was in effect.

For the dependent variable, the variable 'Total cost variable' was used.

II) Judging good fitting regression models and choosing the best regression models

The standards used for judging qualified regression models for this project was based on

- A judgment of the relevance of the independent variables to the variable cost using subject information and common sense.
- A hypothesis test on the overall regression model.
- Standard error of the estimate.
- Adjusted R square value of the regression model.
- 95% of significance level for each independent variables.
- Plots of forecasts and residuals.

The standards used for choosing the best regression models out of the qualified regression models were based on

- Standard error of the estimate.
- Adjusted R square value of the regression model.
- Plots of forecasts and residuals.
- In case the number of variables in the regression models was different from each other, Hocking's criterion for Cp would be used2.

[2] Cp statistics = SSE/MSE - (n - 2p) where n is the number of observations, and p is the number of explanatory variables counting the intercept.
Hocking criterion for Cp is that to choose a regression model as a good one where Cp is equal to or less than the formula, 2p – Total p(=maximum number of parameters of the candidate regression models) + 1. Among these chosen regression models, those with fewer terms are generally considered better regression models.

- Comparing coefficients of the terms.
- Subject information.

The standards for deciding good models was established for the parameter estimation in terms of the dependent variable rather than prediction of the dependent, as this was the primary concern of the project.

III) Implementation

Phase 1

- Check the pattern of the variable cost by time

First, a simple linear regression model between the variable cost and Time was calculated to see if the total variable cost could be explained in the time elapses.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Time[b]	.	Enter

a. Dependent Variable: Total_Cost_Variable

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.593[a]	.352	.331	856.27072

a. Predictors: (Constant), Time

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11957798.81	1	11957798.81	16.309	.000[b]
	Residual	21995986.16	30	733199.539		
	Total	33953784.97	31			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Time

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7645.554	309.975		24.665	.000
	Time	66.207	16.394	.593	4.038	.000

a. Dependent Variable: Total_Cost_Variable

Figure 5 Regression modeling with TVC and Time from SPSS

$$\text{Total cost variable} = 7645.554 + 66.207 * \text{Time}$$

Equation 2 Regression model with TVC and Time

With this regression model, the total variable cost appeared to increase by 66.207 when one month passed. But clearly, the total variable cost did not always increase as time went up, and with R square as seen above, it was not very simple to explain the variability of the dependent variable with only the time factor.

Phase 2

- Using and combining other variables
- Correlation tests

To take other independent variables into account, first some correlation tests had to be performed on the selected variables so that the correlated variables are not in the blurring the regression model.

At this stage, the variables Time, Cust_%, and Production total were selected as predictor variables where Cust_% was meant to control for the factor of product mix.

The correlations tests between these variables showed that the production volume and the percentage of the customer percent are highly correlated and therefore should not be in the same regression model.

	Time	Cust_%	Production_volume
Time		-0.335	0.429
Cust_%			-0.794
Production_volume			

Table 1 Correlation tests

Therefore the following two regression models were obtained.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.747[a]	.559	.528	718.90186

a. Predictors: (Constant), Production_Total, Time

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18966008.20	2	9483004.100	18.349	.000[b]
	Residual	14987776.77	29	516819.889		
	Total	33953784.97	31			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Production_Total, Time

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4745.416	829.447		5.721	.000
	Time	45.006	14.920	.403	3.017	.005
	Production_Total	.173	.047	.492	3.682	.001

a. Dependent Variable: Total_Cost_Variable

Figure 8 Regression modeling with Time, and Prod_volume

Total cost variable = 4745.416 + 0.173*Production_total + 45.006*Time

Equation 3 Regression model with Time, and Prod_volume

1) Time was for understanding the variability of the dependent along time elapses, and Production_total was the one object to be found in terms of the relationship with the variable cost.
2) The hypothesis that this regression model was insignificant was rejected because of the significance level of 0 in the ANOVA table, meaning at least one parameter of the terms in this regression model was not 0, thus making this regression model to be considered.
3) Standard error of the estimate was 718.9
4) The adjusted R square of this regression model was 0.528 which was not too bad nor good.
5) The p values for the independent variables were less than 0.95 and thus were all acceptable.
6) The plot of forecast and residual showed quite good fitting of the model.

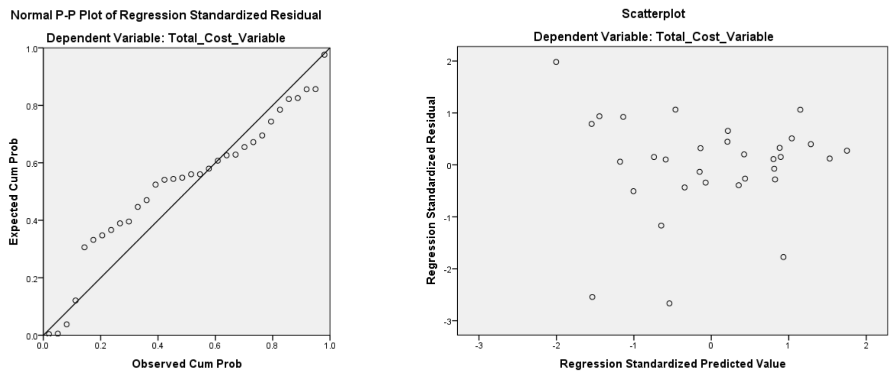


Figure 9 Plots showing the fitness of the model.

Another regression model was as follows.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.640[a]	.410	.369	831.19187

a. Predictors: (Constant), Customer_Percent, Time

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13918267.15	2	6959133.575	10.073	.000[b]
	Residual	20035517.82	29	690879.925		
	Total	33953784.97	31			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Customer_Percent, Time

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12139.205	2684.516		4.522	.000
	Time	59.928	16.345	.537	3.667	.001
	Customer_Percent	-5701.830	3384.820	-.247	-1.685	.103

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 12139.205 + 59.928*Time – 5701.83*Customer %

1) Time was for understanding the variability along time elapses, and Cust_% was for understanding how the shipping pounds to all the customers except Franklin affects the dependent variable.

2) The hypothesis that this regression model was insignificant was rejected because of the significance level of 0 in the ANOVA table, meaning at least one parameter of the terms in this regression model is not 0, thus making this regression model to be considered.
3) Standard error of the estimate was 831.1.
4) The adjusted R square of this regression model was 0.369 which was not good.
5) The p value for Customer_percent was over 0.05 which made it insignificant.

The low adjusted R square of 0.369 and the p value of Customer % higher than 0.05 let this equation unqualified to be considered.

Based on this information, the first regression model had only one added variable, Time compared to the very initial regression equation checked at the very first place and with the adjusted R square of 0.528, it could not be said to explain the dependent variable sufficiently.

Therefore, another factor had to be considered.

Phase 3

- Considering the volatility of the production volume

As mentioned in the introduction section, the variable cost was not a deterministic function of the production volume. As an example of this discordance, the variable cost of a month in which the production volume was 17 million Lbs and when the production volume had been 17million Lbs for the past 6 months as well was not close to the variable cost of a month in which the production volume was also 17 million Lbs, but had been 20 million Lbs for the past 3 months.

Therefore, a variable for the volatilities of the production volume of each month was created and added to the regression model to see the effect the percent changes in the production volume had on the variable cost.

With correlation tests, the variable for volatility was shown to be able to be in a regression model with other variables.

	Time	Cust_%	Production_volume	Volatility by month
Time		-0.335	0.429	0.045
Cust_%			-0.794	0.34
Production_volume				-0.5
Volatility by month				

With regression modeling with these variables,

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.799[a]	.638	.597	649.29572

a. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, Production_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20039426.37	3	6679808.789	15.845	.000[b]
	Residual	11382793.31	27	421584.937		
	Total	31422219.68	30			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, Production_Total

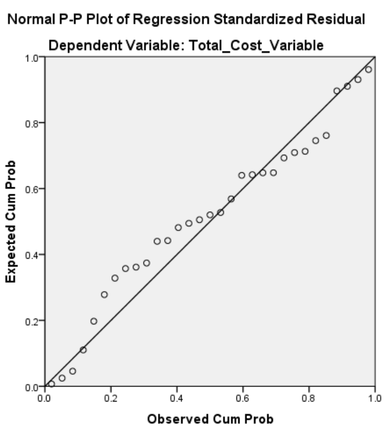
Coefficients[a]

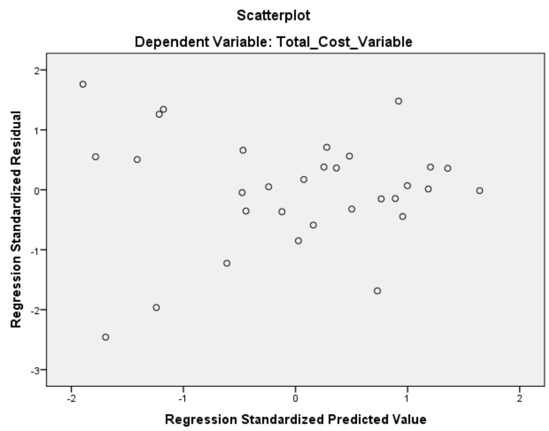
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1247.450	1438.464		.867	.393
	Time	32.973	14.366	.293	2.295	.030
	Production_Total	.253	.051	.726	4.933	.000
	Volatility_of_Production_in_everymonth	2148.370	739.251	.403	2.906	.007

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 1247.45 + 32.973*Time + 0.253*Production_total + 2148.37*Volatility

1) The hypothesis that this regression model was insignificant was rejected because of the significance level of 0 in the ANOVA table, meaning at least one parameter of the terms in this regression model is not 0, thus making this regression model to be considered.
2) Standard error of the estimate was 649.2.
3) The adjusted R square of this regression model was 0.597 which was quite explanatory of the variable cost.
4) The p values for the independent variable were lower than 0.05.
5) The plots of forecast and residuals shows good fitting of the regression model.





With the overall model significance, the p values for each independent variables, and the R square of 0.597, this regression model was considered to be a qualified model for this project.

Modeling with different combinations of the variables all resulted in low adjusted R squares or insignificances of the independent variables, and hence those regression models were discarded.

Phase 4

- Considering the forecast data on shipping Lbs.

Because Aurubis decided how much to produce in a month based on the forecast data on shipping Lbs from 30 to 90 days prior to that month, the forecast shipping Lbs data of 30 days prior to a month were considered to be a variable in the regression model.

Based on correlation tests between variables, the variable of the forecast data should be in a regression model with only Time, or/and Volatility.

	Time	Cust_%	Production_volume	Volatility by month	Forecast
Time		-0.335	0.429	0.045	0.406
Cust_%			-0.794	0.34	-0.87
Production_volume				-0.5	0.85
Volatility by month					-0.41
Forecast					

Regression modeling with all of the combinations of the independent variables with Forecast resulted in low R squares and insignificances, thus those regression models were discarded. Below was the only most fitting model in this stage.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.584[a]	.341	.268	875.63232

a. Predictors: (Constant), TotalShipForecast_30DaysBefore, Time, Volatility_of_Production_in_everymonth

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10720456.80	3	3573485.600	4.661	.009[b]
	Residual	20701762.88	27	766731.958		
	Total	31422219.68	30			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), TotalShipForecast_30DaysBefore, Time, Volatility_of_Production_in_everymonth

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5982.423	1665.277		3.592	.001
	Time	55.062	18.868	.489	2.918	.007
	Volatility_of_Production_in_everymonth	615.129	933.776	.115	.659	.516
	TotalShipForecast_30DaysBefore	.063	.057	.203	1.106	.278

a. Dependent Variable: Total_Cost_Variable

Figure 6 The low R squared and insignificant variables disqualify this model for the project.

Phase 5

- Considering the forecast data involved with production volume.

At this stage, the percentage change in forecast to production volume was considered to be a variable and included in the regression analysis.

With correlation tests, this variable was seen good to be included in a regression model with all of the other variables.

	Time	Cust_%	Production_volume	Volatility by month	Forecast	% chg in FC to Prod
Time		-0.335	0.429	0.045	0.406	-0.033
Cust_%			-0.794	0.34	-0.87	0.224
Production_volume				-0.5	0.85	0.122
Volatility by month					-0.41	-0.082
Forecast						-0.399
% chg in FC to Prod						

Modeling with combinations of the variables, the following two regression models were obtained to be the most optimal ones.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.863[a]	.744	.717	557.12391

a. Predictors: (Constant), ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Time, Production_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25262947.53	3	8420982.509	27.131	.000[b]
	Residual	8690837.443	28	310387.052		
	Total	33953784.97	31			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Time, Production_Total

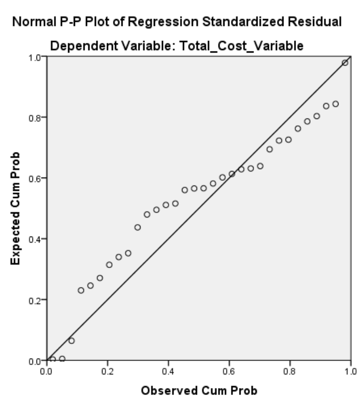
Coefficients[a]

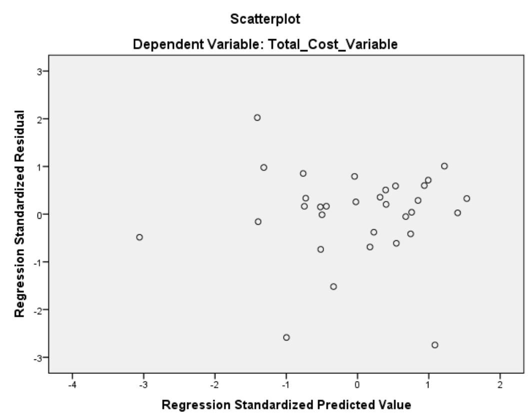
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5257.703	652.777		8.054	.000
	Time	48.532	11.589	.435	4.188	.000
	Production_Total	.148	.037	.421	4.014	.000
	ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal	4094.836	909.125	.436	4.504	.000

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 5257.703 + 48.532*Time + 0.148*Production_total + 4094.836*%chg_in_FC_to_Prod

1) The hypothesis that this regression model was insignificant was rejected because of the significance level of 0 in the ANOVA table, meaning at least one parameter of the terms in this regression model is not 0, thus making this regression model to be considered.
2) Standard error of the estimate was 557.1.
3) The adjusted R square of this regression model was 0.717 which was good to explain the dependent.
4) The p values for all of the independent variable were lower than 0.05.
5) The plot of forecast and residual was acceptable.





Therefore, this model would be useful for this project.

Another regression model was

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.915[a]	.838	.813	442.70896

a. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Production_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26326447.94	4	6581611.985	33.581	.000[b]
	Residual	5095771.738	26	195991.221		
	Total	31422219.68	30			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Production_Total

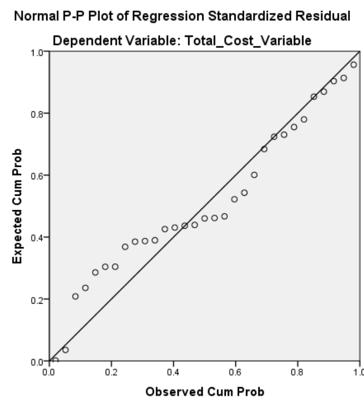
Coefficients[a]

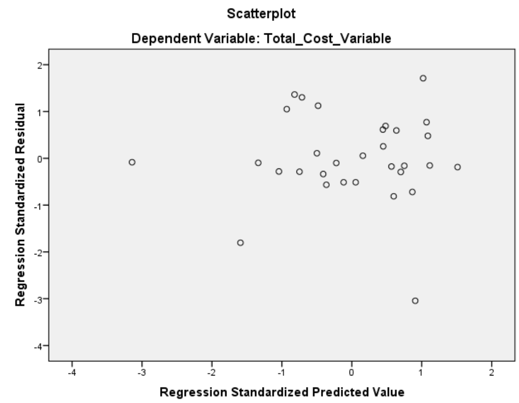
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1758.106	984.922		1.785	.086
	Time	36.575	9.815	.325	3.726	.001
	Production_Total	.228	.035	.654	6.465	.000
	ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal	4091.679	722.433	.452	5.664	.000
	Volatility_of_Production_in_everymonth	2147.170	504.043	.403	4.260	.000

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 1758.106 + 36.575*Time + 0.228*Production_total + 4091.679*%chg_in_FC_to_Prod + 2147.17*Volatility

1) The hypothesis that this regression model was insignificant was rejected because of the significance level of 0 in the ANOVA table, meaning at least one parameter of the terms in this regression model is not 0, thus making this regression model to be considered.
2) Standard error of the estimate was 442.7.
3) The adjusted R square of this regression model was 0.813 which was even better to explain the dependent.
4) The p values for all of the independent variables were lower than 0.05.
5) The plot of forecast and residual was good.



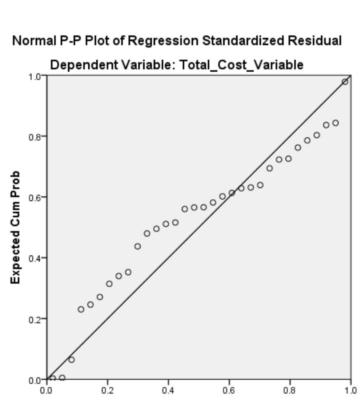


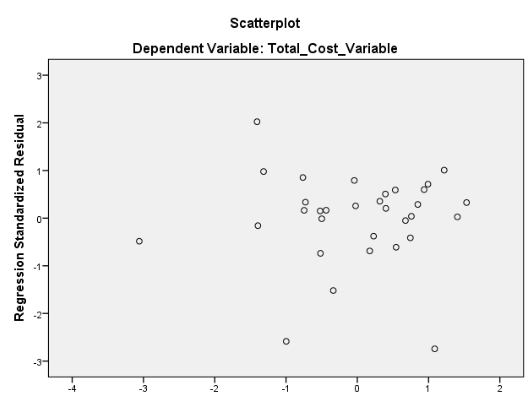
This model was qualified for the project.

Comparing both the regression models above, the first one had

1) Standard error of the estimate 557.1.
2) 0.717 R squared.
3) Hocking criterion for Cp is acceptable.

   Cp = (8690837.443/310387.052) - (32 - 2*4) = 3.999999958 < 2*4 - 5 + 1 = 4
4) Plot was acceptable, but a little less fit than the second regression model.

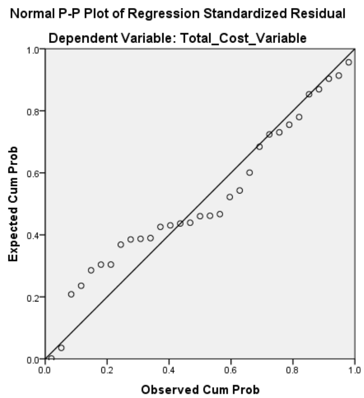


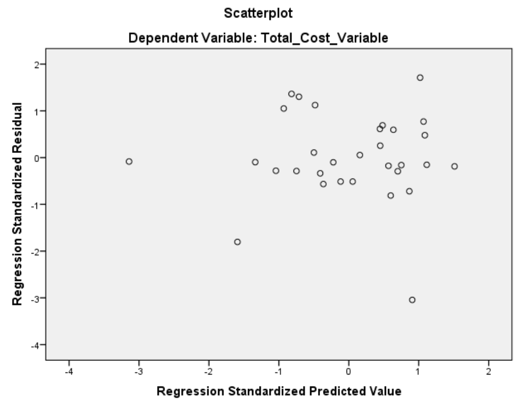


The second model had

1) Standard error of the estimate 442.7.
2) 0.813 R squared.
3) Hocking criterion for Cp is acceptable.

   Cp = (5095771.738/195991.221) - (32 - 2*5) = 3.999999959 < 10 - 5 + 1 = 6
4) Plot was slightly more fit than the first regression model.





When the coefficients of the terms were compared, the first regression model had a bit too deviant coefficients compared to the other regression model in this stage and the other qualified model in phase 3, and might not make sense in consideration of common sense and subject information.

Therefore, based on the standard error of estimate, the adjusted R square, fitness of the plots, and the interpretation of the coefficients, the second model in this stage was chosen as the one qualified model in this stage.

To select the best regression model for this project, the two qualified regression models obtained in the research procedures are compared.

[Regression model A – obtained in phase 3]

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.799[a]	.638	.597	649.29572

a. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, Production_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20039426.37	3	6679808.789	15.845	.000[b]
	Residual	11382793.31	27	421584.937		
	Total	31422219.68	30			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, Production_Total

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1247.450	1438.464		.867	.393
	Time	32.973	14.366	.293	2.295	.030
	Production_Total	.253	.051	.726	4.933	.000
	Volatility_of_Production_in_everymonth	2148.370	739.251	.403	2.906	.007

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 1247.45 + 32.973*Time + 0.253*Production_total + 2148.37*Volatility

[Regression model B – obtained in phase 5]

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.915[a]	.838	.813	442.70896

a. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Production_Total

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26326447.94	4	6581611.985	33.581	.000[b]
	Residual	5095771.738	26	195991.221		
	Total	31422219.68	30			

a. Dependent Variable: Total_Cost_Variable

b. Predictors: (Constant), Volatility_of_Production_in_everymonth, Time, ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal, Production_Total

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1758.106	984.922		1.785	.086
	Time	36.575	9.815	.325	3.726	.001
	Production_Total	.228	.035	.654	6.465	.000
	ProductionTotal_minus_Forecast__ThenDividedBy_ProdTotal	4091.679	722.433	.452	5.664	.000
	Volatility_of_Production_in_everymonth	2147.170	504.043	.403	4.260	.000

a. Dependent Variable: Total_Cost_Variable

Total cost variable = 1758.106 + 36.575*Time + 0.228*Production_total + 4091.679*%chg_in_FC_to_Prod + 2147.17*Volatility

1) Standard error of the estimate
    - Regression model A = 649.2
    - Regression model B = 442.7

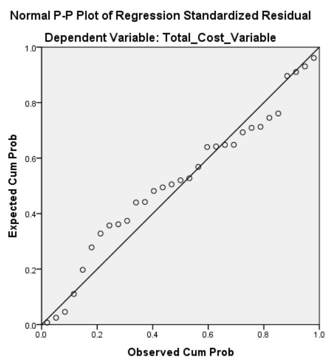
    Regression model B had less errors.

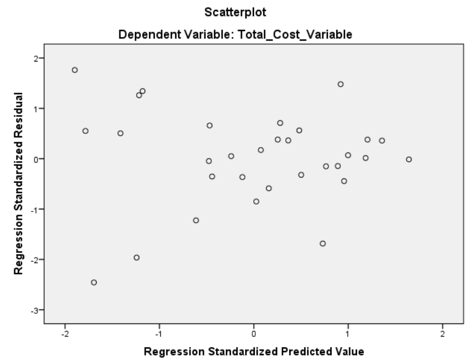
2) Adjusted R square
    - Regression model A = 0.597
    - Regression model B = 0.813

    B had higher explanation of the dependent.

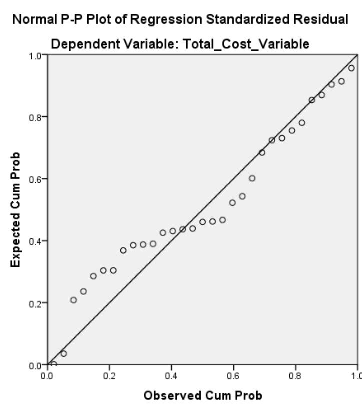
3) Plot

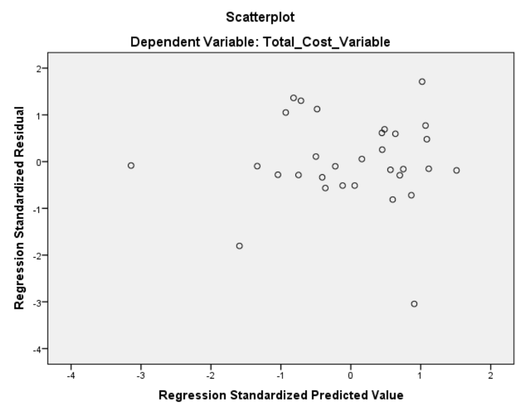
- Regression model A





- Regression model B





Regression model B had slightly better fit than the regression model A.

4) Hocking criterion for Cp

- Regression model A : Acceptable.

  Cp = (11382793.31/421584.937) - (32 - 2*4) = 3.000000026 < 2*4 - 5 + 1 = 4

- Regression model B : Acceptable.

  Cp = (5095771.738/195991.221) - (32 - 2*5) = 3.999999959 < 10 - 5 + 1 = 6

5) Coefficients

Both models had similar coefficients for each corresponding terms, thus making sense.

6) Subject information

With all of the variables A had, B had one more variable of the percentage change in forecast to production volume. Because Aurubis set the producing amount of a month prior to that month based on the forecast data on the demands, the budget, labor fees and etc had been fixed, and it could lead to waste of resources if the expected demands and the actual shipping pounds are much different. In this context, it is more desirable to include the percentage change in forecast to production volume as an exploratory variable on top of the variables the model B had.

Overall, based on the standard error of the estimate, the adjusted R square, fitness from the plots, and the aspects of subject knowledge, the regression model B is better to explain the variable cost and thus adopted for this project.

Total cost variable = 1758.106 + 36.575*Time + 0.228*Production_total + 4091.679*%chg_in_FC_to_Prod + 2147.17*Volatility