ISSN: 04532198
Volume 62, Issue 04, April, 2020

Unfolding the Message from the City Gas Sector in Indonesia

Andry Prima[1,*], Astri Rinanti[2], Onnie Ridaliani[3], Wiwik Dahani[4], Mustamina Maulani[5], Abdul Hamid[6], Bayu Satiyawira[7]

Petroleum Department, Universitas Trisakti, DKI 11440, Indonesia[1,3,5,6,7]
Environmental Engineering Department, Universitas Trisakti, DKI 11440, Indonesia[2]
Mining Department, Universitas Trisakti, DKI 11440,Indonesia[4]

Corresponding Author: 1*

Abstract— One of the milestones any company or industry should consistently achieve is the long-term steady growth. That nonetheless is precisely the position of natural gas distribution sector-without exception. Year-over-year statistical data should portray the broader picture of city gas sector in Indonesia that can possibly be explained by considering a number of key elements such as the workforces involved in the industry, wages that have to be spent on the labours and finally yet importantly volume of natural gas distributed for the city consumers. Throughout this particular study, a set of variables taken into account and then subsequently perform the analysis by using multivariate regression analysis as the selected method. The purpose of this particular study is to explain a better understanding of Indonesia's city gas distribution sector. The methodology used in this study is the multivariate linear regression. In general, this paper is directed toward defining the notion of the growth of natural gas distribution sector in Indonesia. At the end of this study, the linear model effectively explains the strong correlation between the revenue (dependent variable) of the city gas distribution sector with other independent variables predictors) and fits to represent the macro perspective of city gas sector in Indonesia. In that respect, by incorporating a number of variables as the input of regression analysis tools, the expected result is a model that can be applied to predict the long-term growth.

Keywords— City Gas, Explanatory Variables, Gas Distribution, Linear Model, Regression Analysis, Revenue, Workforce

1. Introduction

One of the milestones any company or industry should consistently achieve is the long-term steady growth. That nonetheless is precisely the position of natural gas distribution sector-without exception. Year-over-year statistical data should portray the broader picture of city gas sector [1] in Indonesia that can possibly be explained by considering a number of key elements such as the workforces involved in the industry, wages that have to be spent on the labours and finally yet importantly volume of natural gas distributed for the city consumers. The purpose of this particular study is to explain a better understanding of Indonesia's city gas distribution sector and to provide a clearer picture of a number of explanatory variables should be considered to optimize the distribution of natural gas to cities in Indonesia. The methodology used in this study is the multivariate linear regression [2]. In more specific, the explanatory multivariate regression analysis is used to analyze correlations between variables and establishing the validity of the multiple regression model. In this regard, considering the following data set released from the Indonesia Central Bureau of Statistics that collected the data of city gas distribution company during years 2010-2017 consisting of: revenues, number of workforces, wages, volume of city gas distributed, and cost associated with operations [3]. Findings - a chance to give suggestion to government and other stakeholders in terms of

optimizing the growth of Indonesia's gas distribution sector. The value or the originality of this study is to offer recommendation for optimization based on the hierarchy of explanatory variables; from the top priority to least important one. Study limitations - Due to the inaccessible primary non-disclosure data and the limited data availability, it is assumed that the data derived from Indonesia Central Bureau of Statistics are valid for describing the explanatory variables of Indonesia's city gas distribution sector. Several information from a data set collected from Indonesia Central Bureau of Statistics [4]. The following figures (Figures 1 to 4) explains the information related to city gas distribution sector in Indonesia.

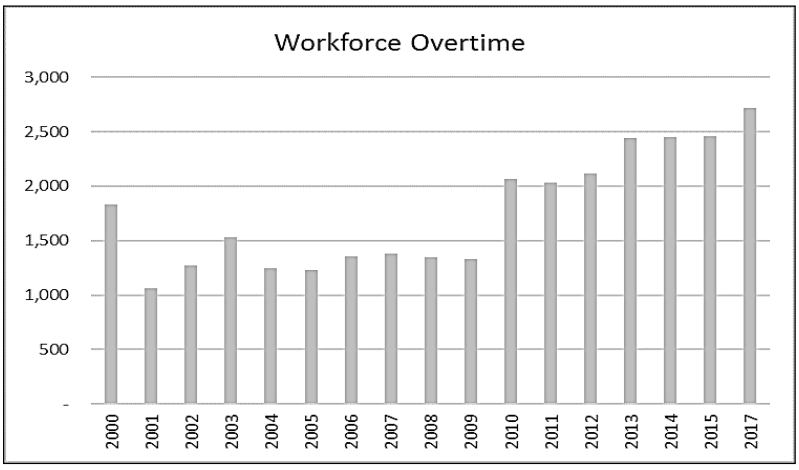

Figure 1. Workforce vs time

Figure 2. Wages vs time

ISSN: 04532198
Volume 62, Issue 04, April, 2020

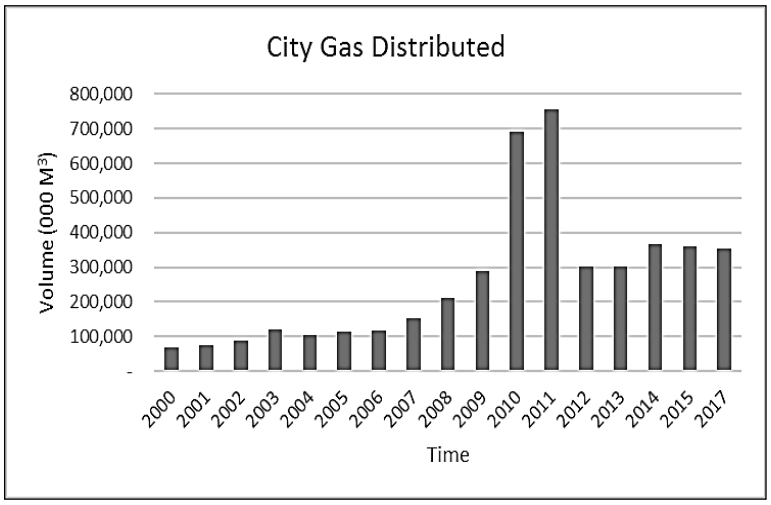

Figure 3. City gas distributed vs time

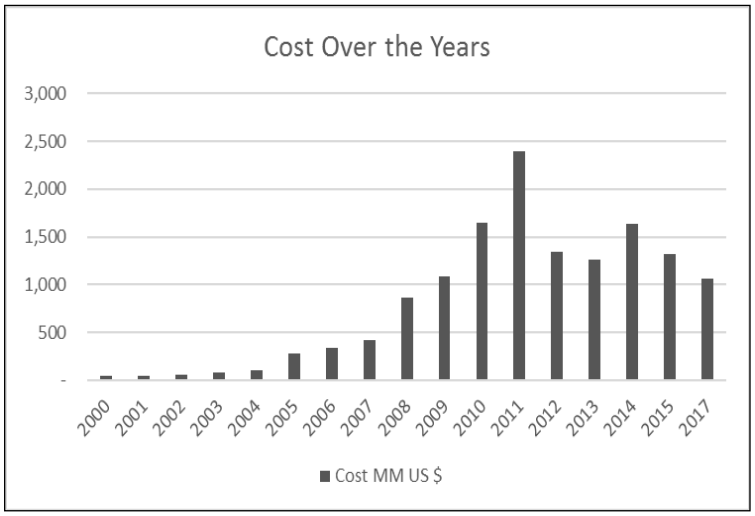

Figure 4. Cost vs time

2. The Method

In this section of this study, the method of how to process the data set is described. The selected method appears to be the linear regression. In general, the utilization of linear regression model is useful to explain certain reasons, [6], [7], [8], [9]. The samples taken during year of 2000 to 2017. The samples collected from the resources of information is aggregate total within the Industry.

a. The description of variables - It can be used to analyze the strength of the association between the outcome (dependent variable) and predictor variables.
b. Adjustment - It adjusts for the effect of covariates or the confounders, K and Yadav S 2019 [10].

In more specific, the explanatory multivariate method multiple regression analysis is used to analyze correlations between variables and establishing the validity of the multiple regression model. In this regard, considering the following data set released from the Indonesia Central Bureau of Statistics that collected the data of city gas distribution company during years 2010-2017 [11], [12], [13] consisting of: revenues, number of workforces, wages, volume of city gas distributed, and cost associated with operations. The framework built in their study incorporates variables that were treated as dependent variable that at the next step would be influenced by the free variables. The research model refers to previous study by Paikun,

T.Kadri, , & R.D. Sugara [2], [14]. The model proposed in the said study is geared toward linear relationship between dependent and dependent model as shown Figure 5 [2], [15].

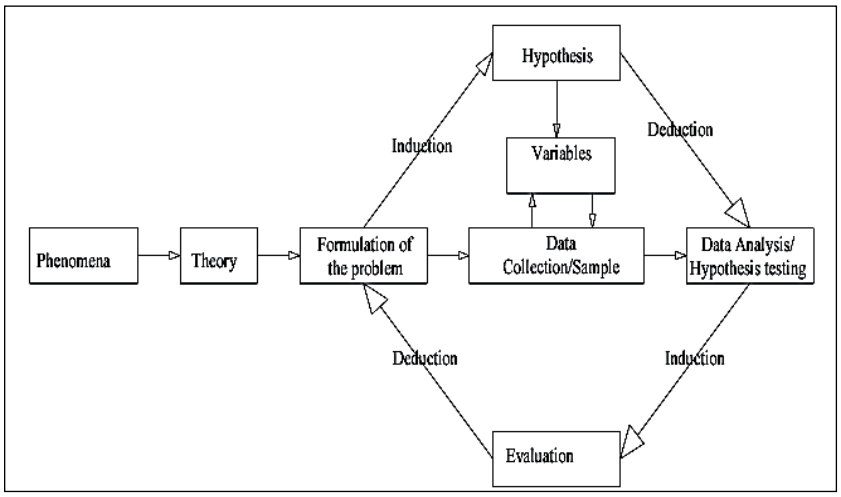

Figure 5. Linear model

More in-depth analysis can be done through the Regression approach [16], [17]. From the first table tabulation above, the Non Linear pattern (polynomial) should be used, with the basic equation pattern being:

$Y = am1x1m+\ldots+amnxnm+\ldots+a11x1+\ldots+a1nxn+b$ ..... (1)

In order for the model fits the data set, certain parameters must be taken into account:

2.1. R2 as Coefficient of Determination

R2 is statistically the coefficient of determination that is the portion of the total variation in the dependent variable that can be explained by variation in the independent variable(s). When R2 is + 1, there exists a perfect linear relationship between x and y, i.e., 100% of the variation in y is explained by variation in x. When it is 0< R2 <1, there is a weaker linear relationship between x and y, i.e., some, but not all of the variation in y is explained by variation in x [18], [19].

2.2. Residuals Plot

A residual plot normally shows a graph in which the residuals are plotted on the vertical axis, whereas the independent variable on the horizontal axis. As far as the graph, if the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is highly likely appropriate for the data. Otherwise, a non-linear model should be opted as it is more appropriate [20], [21].

3. The Discussions

The following graphics are the explanatory figures presented and analyzed in order to explain the correlation between the dependent variable (in this particular study is the Revenue) and several other variables such as workforce, wages, volume of gas distributed and cost as the purpose of this particular study is to explain a better understanding of Indonesia's city gas distribution sector and the following 3.1 and 3.2 provide a clearer picture of a number of explanatory variables (see Table 1).

Table 1. Model descriptive statistics

	Mean	Std. Deviation	N
Revenue	2.101.65	1.961.162	17

Workforce	1.757.82	537.814	17
Wages	29.81	18.454	17
Distributed	263.119.76	204.132.765	17
Cost	824.38	715.210	17

Furthermore, the discussion should have discovered that the linear model was the fit one for this data set [8], [9], [10]. The research results of this study starts from the following Table 1 showed the description of the model and Table 2 summarized the model developed using SPSS multivariate linear regression features [11], [12], [13]. Commencing the model summary, the description of the model was automatically generated [14], [15], [16], [17]. The Table 1 provided the general information related to the model, which then later could presumably help determine the dominant factors in the model.

Table 2. Model summary model

Model R	R Square	Adjusted R Square
.989	.979	.972

In order for the model fitted the data set, certain parameters must have been taken into account [18], [19]:

3.1. Overall Model Fit

As shown in the table 2 above, the R-Square that was statistically the proportion of variance (as it values might have varied) in the dependent variable (in this special case Revenue), which could predict a number of independent variables; workforce, wages, the volume of gas distributed into the city, and the cost associated with the distribution of city gas respectively. The value of R-Square statistically indicated that 97.20% of the variance in Revenue could be predicted from the previously mentioned independent variables. Here, 97.20% points fell within the line. The adjusted R-square further strengthened the model implicitly. If all predictors were incorporated to the model, each predictor would then explain the possibility of developing different scenario due to the change of the variance of the dependent variables. Thus, the increase or decrease in R-square would be simply due to chance variation in this particular study. For instance, when this study was conducted, the value of R-square was 97.20%. Therefore, the model was theoretically fit and could presumably be applicable in this study. Adjusted R-square adjusted effectively for multiple variables.

3.2. Residuals Plot

A residual plot normally shows a graph in which the residuals are plotted on the vertical axis, whereas the independent variable on the horizontal axis. As far as the graph, if the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is highly likely appropriate for the data. Otherwise, non-linear model should be opted as it is more appropriate [20], [21] (Figure 6).

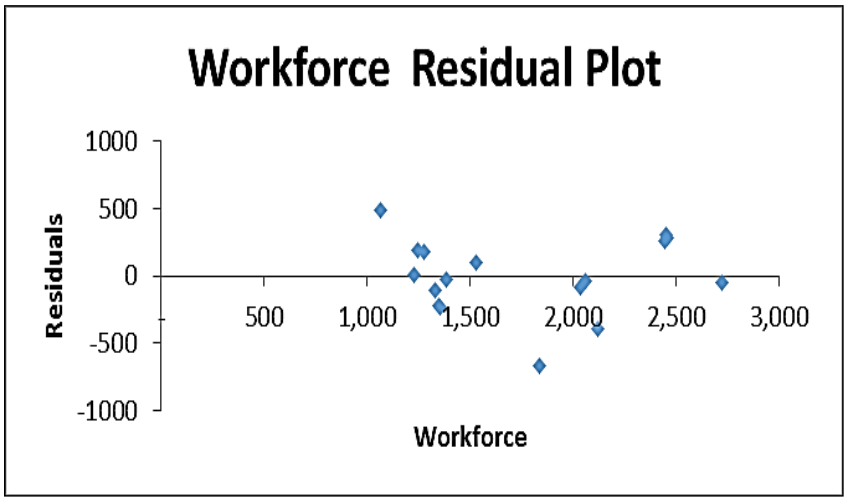

Figure 6. Workforce residual plot – linear model

The linear model developed in this study incorporates more than one explanatory variable. A number of tables below shows inputs and outputs from the linear regression analysis. With regard to the aim of this study, in order to provide a clearer picture of a number of explanatory variables, a model is needed. As the model develop a linear pattern, as seen in Figure 7, the aim of this study is effectively obtained [22], [23].

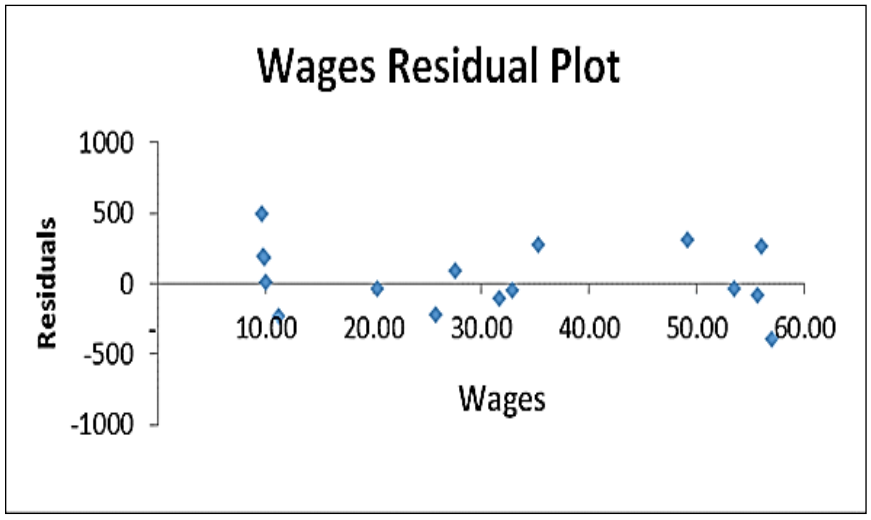

Figure 7. Wages residual plot – linear model

As seen from Figure 6 to 10, points of data form randomly dispersed pattern around the horizontal signifying the message that the model developed in this particular study theoretically fits to represent the macro perspective of city gas sector in Indonesia. For instance, the workforce (Figure 6) residual plot visualized trend of linear model. The work force data form a random pattern indicated that workforce as a variable fitted in this model. Each single data moved along the axis data randomly developing pattern that was only fit for linear model.

ISSN: 04532198
Volume 62, Issue 04, April, 2020

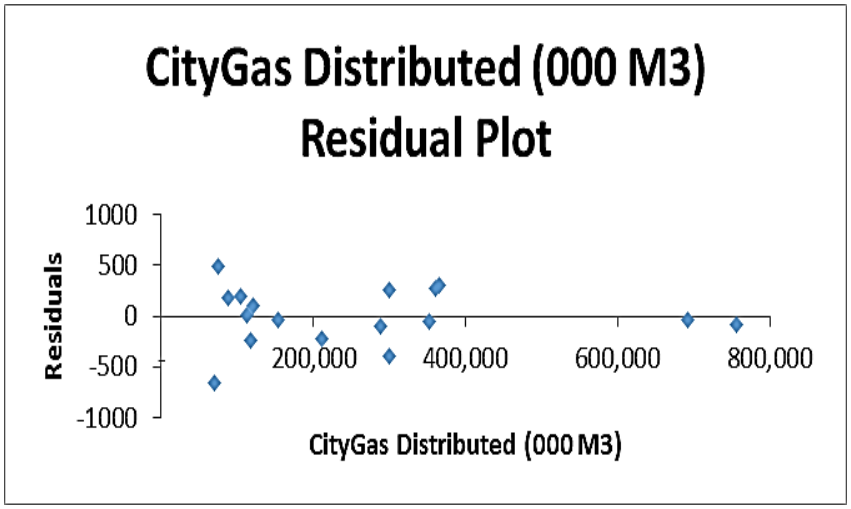

Figure 8. City gas distributed residual plot – linear model

Similarly, wages residual plot resembled that of random pattern's. The wages data form a uniquely outlined that matched the set of data normally formed in a linear model. Likewise, the volume of city gas distributed into the city residual was indifferent that of the random pattern's. The city gas distributed data appeared to be scattered set of data along the horizontal x-axis. From Figure 9, the line fit plot could technically be used to explain the relationship between one continuous predictor or the predicted revenue in this case and a response of the revenue. Furthermore, the relationship between the data of revenue and the predicted value were clear to help better understand that both predictor and response moved proportionally.

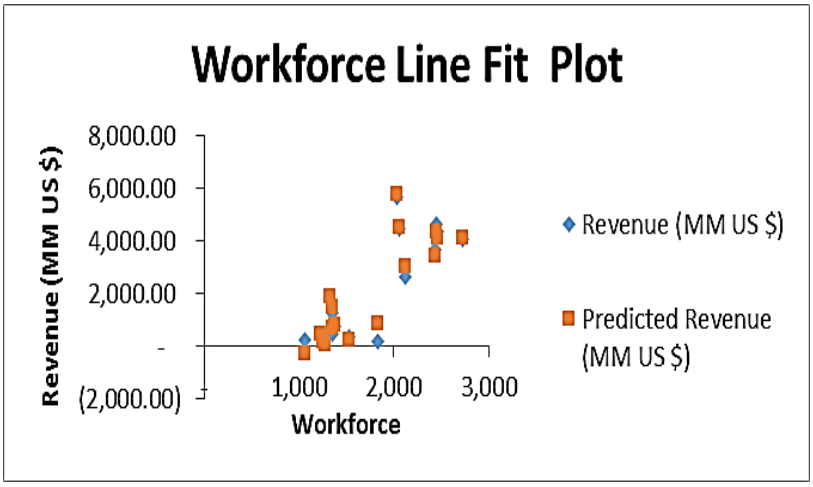

Figure 9. Workforce line fit plot – linear model

From Figure 10, the line fit plot suggested that the relationship between the data of city gas distributed and its predictor were useful to explain a solid linear relationship between the two.

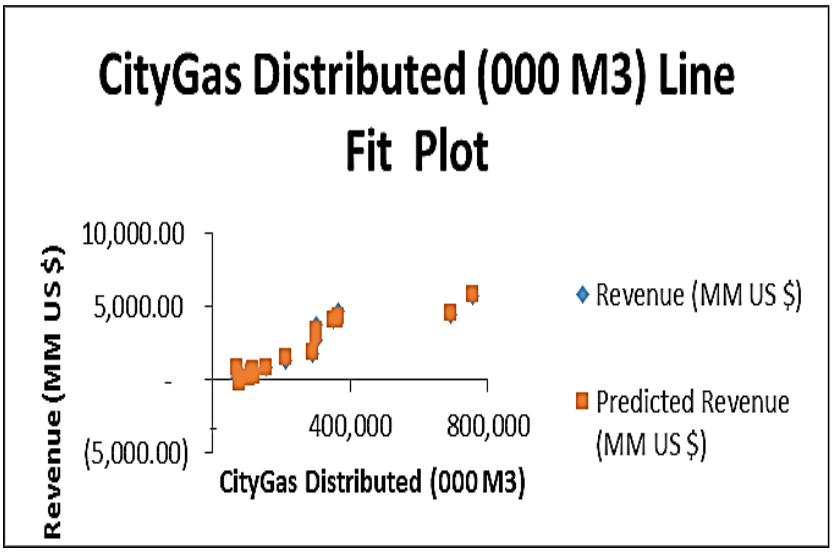

Figure 10. City gas line fit plot – linear model

In this particular study, the data were plotted using regression function of MSExcel and then generated a normal probability plot for the model being developed, Figure 11. In addition, the auto-generated graph produced a nearly linear trend.

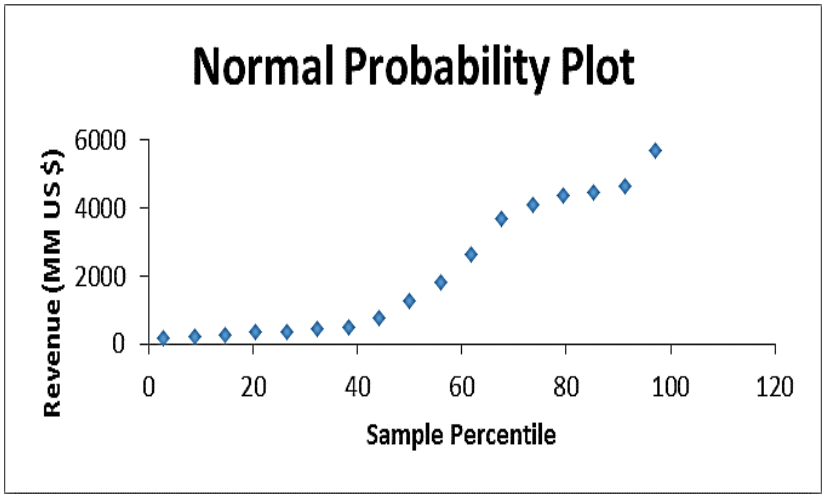

Figure 11. Model normal probability plot – linear model

As a result, a strong signal good model for this data set was finally discovered due to the normal distribution shape. After collecting the data and then put it in the data processing tools, in this particular MSExcel, by using regression, one its features, then it produced a normal probability plot for the model. From Figure 12, the standardized residuals came into sight to be normally distribution. In one side of the plotted data, the graph might have looked slightly off-centre. Nevertheless, the whole pattern was not extremely deviated from being an ideal and normal distribution. Thus, no violations of the normality assumption was found.

ISSN: 04532198
Volume 62, Issue 04, April, 2020

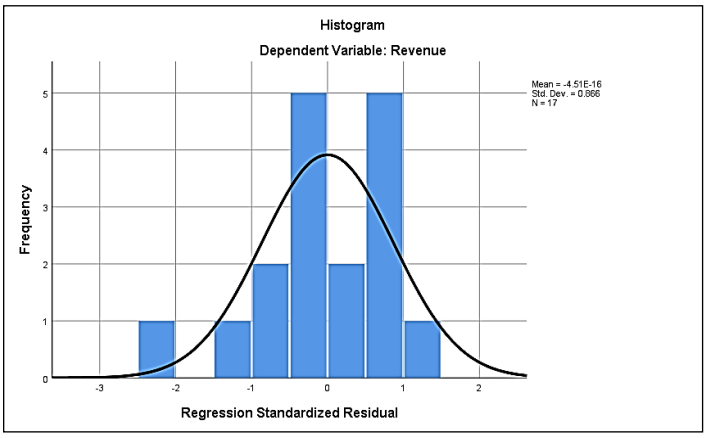

Figure 12. The histogram of target revenue

The Figure 13 displayed that the graph looked slightly skewed. However, such image could not be hugely deviated from being a normal distribution. Accordingly, from this figure, the shape of distribution satisfied the normality assumption. Furthermore, the pattern exhibited there was an explicitly strong indication that the residuals were normally distributed at each level of Y and constant in variance across levels of Y. Therefore, these residuals appeared to be approximately normally distributed.

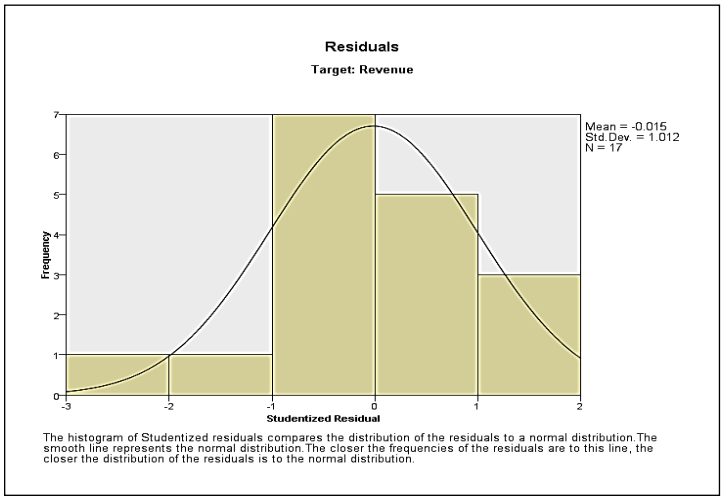

Figure 13. The histogram of studentized residual.

Figure 14 ranked each input field based on the strength of its relationship to the specified target, independent of other inputs (variables) and indicated the relative importance of each variable for this particular linear model. Surprisingly, both tables showed that most important predictor was the workforce followed by cost, distributed gas and wages respectively.

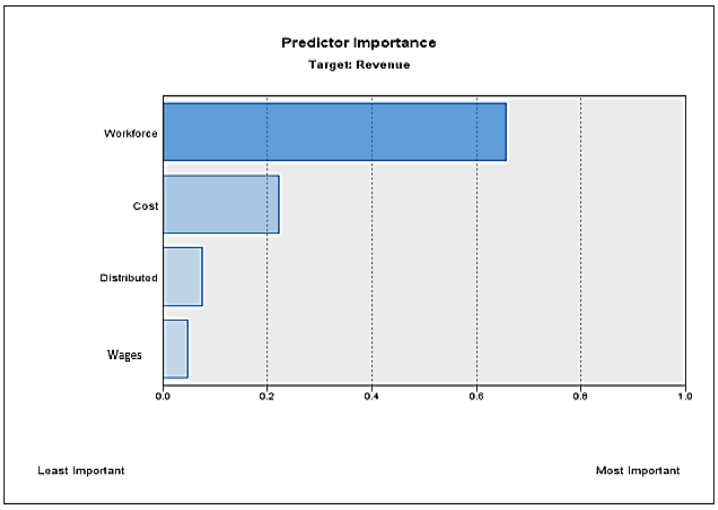

Figure 14. Predictor importance

Finally, this study revealed that the linear model effectively explained that the interaction between variables did exist. Again, the model, as seen from Figure 15, was statistically very accurate. In more details, the observed values of a single numeric variable, in this particular case the revenue, were plotted against the expected values. Then result of the predicted revenue by observe, the sample formed a normal distribution, as all points clustered around a straight line. Therefore, the strong correlation between the model's predictions and its actual results did exist.

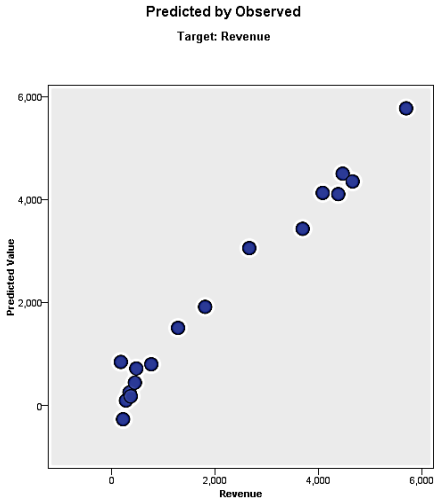

Figure 15. Predicted revenue by observed

4. Conclusions

At the end of this study, the linear model effectively explains the strong correlation between the revenue (dependent variable) of the city gas distribution sector with other independent variables predictors) and fits to represent the macro perspective of city gas sector in Indonesia. such fiunding Is Important for study since the purpose of this particular study is to explain a better understanding of Indonesia's city gas distribution sector and the following 3.1 and to provide a clearer picture of a number of explanatory variables. For instance, statistically indicates that 97.20% of the variance in Revenue can be predicted from the previously mentioned independent variables such as workforce, wages, volume of gas distributed and cost. The line fit

plot also can technically be used to explain the relationship between one continuous predictor and the revenue. In addition, the auto-generated normal probability graph produced a nearly linear trend. Next, as an explanatory variable, the workface plays most important variable in this model followed by cost, distributed gas and wages respectively. Last but not least, the result of the predicted revenue by observe, the sample formed a normal distribution, as all points clustered around a straight line. Therefore, the linear model is statistically the best fit for the data set of city gas distribution sector in Indonesia.

5. Acknowledgments

Authors wishing to acknowledge assistance or encouragement from all colleagues for their contributory genuine and yet professional thoughts including special technical support by the centre of research of Universitas Trisakti. Special thanks to IBM's SPSS and Microsoft Excel's Analysis tool pack for such helpul and powerful Statistical tools. Last but not least everyone in Universitas Trisakti for their unforgettable and meaningful helps.

6. References

[1] G. K. Uyanık, & N. Güler, "A Study on Multiple Linear Regression Analysis", Procedia - Social and Behavioral Sciences. Volume 106, pp. 234–240, 2013. 10.1016/j.sbspro.2013.12.027.

[2] Paikun, T.Kadri, & R.D. Sugara, "Estimated budget construction housing using linear regression model easy and fast solutions accurate", Int. Conf. on Computing, Engineering, and Design (ICCED), pp. 1-6, 2017, DOI:10.1109/ced.2017.8308095

[3] Indonesia Central Bureau of Statistics, available at: https://www.bps.go.id/dynamictable/2019/03/13/1595/volume-penjualan-gas-alam-melalui-saluran-pipa-menurut-jenis-pelanggan-mmscf-2010-2017.html

[4] Al-Khazraji, A. Khalil & T. Muhannad, "Multiple Linear Regression Approach for the Permeability Calculation from Well Logs: A Case Study in Nahr Umr formation –Subba Oil Field, Iraq", International Journal of Science and Research, 358, 2014, (IJSR) ISSN (Online): 2319-7064

[5] H. Elsayir, "Residual Analysis for Auto-Correlated Econometric Model", Open Journal of Statistics, Volume 09, pp. 48-61, 2019, Doi: 10.4236/ojs.2019.91005.

[6] K. Lay & A.Cho, "Applied SPSS for Business and Marketing", International Journal of Trend in Scientific Research and Development. Volume 3, 2019, Doi:1148-1150. 10.31142/ijtsrd24013.

[7] A. Hutagalung, D. Hartono, M. Arentsen, and J. Lovett, "Economic Implications of Domestic Natural Gas Allocation in Indonesia", International Journal of Energy Sector Management, Volume 13, Number 2, pp. 424-449, 2019. https://doi.org/10.1108/IJESM-05-2018-0003

[8] N. P. Salsabila, W.W. Purwanto, and M.R. Fuad, "Development of Integrated Renewable Energy System Model for Rural Productivity Zone in East Nusa Tenggara", IOP Conference Series: Materials Science and Engineering, 2019, Doi:10.1088/1757-899x/543/1/012070.

[9] W.J. Niu, Z.K. Feng, B.F. Feng, Y.W. Min, C.T. Cheng, J.Z. Zhou, "Comparison of Multiple Linear Regression, Artificial Neural Network, Extreme Learning Machine, and Support Vector Machine in Deriving Operation Rule Of Hydropower Reservoir", Water, 2019, Doi: 11. 88. 10.3390/w11010088.

[10] K. Çelik, "Predicting Chlorophyll-a Concentrations in Two Temperate Reservoirs with Different Trophic States Using Principal Component Regression (PCR)", Oceanological and Hydrobiological Studies, 2018, Doi: 47. 10.1515/ohs-2018-0001.

[11] K Kumari, S Yadav, "Linear regression analysis study", J. Pract Cardiovasc Sci;4:33-6, 2018. DOI: 10.4103/jpcs.jpcs_8_18.

[12] W.W. Purwanto, Y. Muharam, Y.W. Pratama, D. Hartono, H. Soedirman, & R.Anindhito, "Status and Outlook of Natural Gas Industry Development in Indonesia", Journal of Natural Gas Science and Engineering, Volume 29, pp. 55–65, 2016, Doi:10.1016/j.jngse.2015.12.053

[13] D. Hakam, & A. Asekomeh, "Gas Monetisation Intricacies: Evidence from Indonesia", International Journal of Energy Economics and Policy, Volume 8, Number 2, pp. 174-181, 2018, ISSN: 2146-455

[14] L. Endriana, D. Hartono, and T. Irawan, "Green Economy Priority Sectors In Indonesia: A SAM Approach", Environmental Economics and Policy Studies,Volume 18, Number 1, pp.115-135, 2016, Doi:10.1007/ s10018-015-0114-5.

[15] Wang L, Tao G, Wang Z, "Application Research of Data Mining on Reservoir Characterization", 2011 Int. Conf. on Electronic and Mechanical Engineering and Information Technology, pp. 153-156, 2011, 10.1109/EMEIT.2011.6022885, DOI: 10.1109/EMEIT.2011.6022885

[16] J.M. Chambers, W.S. Cleveland, B. Kleiner, P.A, Tukey: Graphical Methods for Data Analysis. Wadsworth & Brooks/Cole, Belmont, CA.1983.

[17] F. Fechete and A. Nedelcu, "Analysis of The Economic Performance of an Organization using Multiple Regression", Proc. Int. Conf. on Scientific Paper Afases (Brasov), 2014.

[18] NIST/SEMATECH e-Handbook of Statistical Method. The National Institute of Standards and Technology (NIST), a physical sciences laboratory, and a non-regulatory agency of the United States Department of Commerce, 2012 available at https://www.itl.nist.gov/div898/handbook/index.htm

[19] F. Liang, Ryvak, S.M. Sayeed, et al., "The Role of Natural Gas as a Primary Fuel in the Near Future, Including Comparisons of Acquisition, Transmission and Waste Handling Costs of as With Competitive Alternatives", Chemistry Central Journal, Volume 6, 2012, Doi:10.1186/1752-153X-6-S1-S4

[20] A. Griffith, 2010, SPSS for Dummies, 2nd edition, Wiley Publishing, Inc 12, p 240- 292, ISBN: 978-0-470-48764-8

[21] R Ho Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS Chapman & Hall/CRC Taylor & Francis Group, 2006, ISBN 1-58488-602-1 (alk. paper)

[22] P. Tharakan, "Summary of Indonesia's Energy Sector Assessment", Asian Development Bank, 2015, Available at: www.adb.org; openaccess.adb.org.

[23] S. Saffarzadeh & S. Shadizadeh, "Reservoir Rock Permeability Prediction using Support Vector Regression in an Iranian Oil Field", Journal of Geophysics and Engineering – J. Geophys Eng, Volume 9, pp. 336-344, 2012, 10.1088/1742-2132/9/3/336.