Development of Welded Line Pipe Price Prediction Model: Comparison Between Linear Regression, ARIMA and ARIMAX Methods at PT Hulu Nasional

Mutiara Sari Farida Hutabarat[1,*] and Nur Aini Masruroh[1]

[1] Mechanical & Industrial Engineering Department, Universitas Gadjah Mada, Yogyakarta, Indonesia
* mutiarasarifaridahut@mail.ugm.ac.id

Abstract. PT Hulu Nasional, a KKKS that operates under a Cost Recovery scheme, is obligated to formulate a material/service procurement cost estimate (Owner Estimate/OE) that is both valid and reasonable, in accordance with the prevailing market price. The material under scrutiny in this study is Welded Line Pipe. One approach to obtaining a reasonable estimate of the price of welded line pipe is to develop a price prediction model based on historical pipe purchase data and the identification of variables that affect the current pipe price, according to market conditions. In this study, a price prediction model was developed, incorporating three distinct approaches: a causal approach (linear regression), a time series approach (ARIMA), and a combination of causal and time series approaches (ARIMAX). The most suitable model is then determined based on its capacity to predict actual data. Preliminary findings from the linear regression analysis model, which has an adjusted R2 value of 0.7, indicate the presence of variables that exert an influence on the cost of pipes. These variables include the price of crude oil (Brent oil), the exchange rate of the U.S. dollar in relation to the Indonesian rupiah, and the geographical location of material delivery to the KTI Region, which encompasses eastern Indonesia. The ARIMAX (1,1,1) method with two exogenous variables, namely crude oil prices (Brent oil) and the dollar exchange rate, emerged as the optimal model based on the lowest RMSE (109.5), MAPE (2.95 %), and MAE (67.35) values when compared to the ARIMA (1,1,1) model and linear regression analysis.

Keywords: Procurement, Regression Linear, ARIMA, ARIMAX, Welded Line Pipe, Owner Estimate.

1 Introduction

PT Hulu Nasional ("PT HN") is a Cooperation Contractor (KKKS) with a cost recovery scheme, which is a form of partnership that allows contractors to recover operating costs after successful oil and gas production [1]. In this scheme, once production starts, the profits earned are first used to reimburse all operational expenses previously incurred by the KKKS. Consequently, contractors are obligated to account for all financial resources, inclusive of those allocated for exploration and production

M. Rafieiy et al. (eds.), *Proceedings of the 8th Mechanical and Industrial Engineering Symposium (MIE 2025)*,
Atlantis Highlights in Engineering 42,
https://doi.org/10.2991/978-94-6239-687-6_28

materials and services. The form of KKKS's responsibility for the operational costs it incurs is by carefully, accurately, and validly estimating procurement costs in accordance with prevailing market prices.

SKK Migas regulated nine primary commodity categories, Line Pipe is among the top 2 commodities with the highest procurement value at 38 % [2]. Line pipe is a strategic material at PT HN due to its substantial value, necessitating meticulous calculation in every procurement process. The primary raw material employed in the fabrication of welded line pipe is Hot Rolled Coil (HRC), accounting for up to 82.5 % of production costs [3]. In the oil and gas industry, two primary types of line pipe are identified: seamless and welded. Presently, welded line pipe is produced domestically, while seamless pipe is imported. Consequently, the present study will concentrate on the analysis of the cost of domestic welded line pipe.

In accordance with SKK Migas regulations, the procurement of materials and services in Cooperation Contractor with a cost recovery scheme is required to undergo an auction process, commencing with the preparation of a price estimate or Owner Estimate (OE) to serve as a reference in determining the bidding. PT HN is obligated to meticulously prepare the OE in a manner that aligns with prevailing market conditions. The objective of this study is to examine the OE for welded line pipe, a critical component due to the substantial value of the procurement. An OE that is excessively high can potentially inflict harm upon the company. Conversely, an OE that is inadequately low has the capacity to impede the auction process and delays material supply, potentially disrupting PT HN's operations.

In order to obtain a reasonable price estimate, this study builds a prediction model based on historical data and current market variables. To do so, it uses Linear Regression, ARIMA, and ARIMAX approaches. The objectives of this study are threefold: first, to analyse the variables that affect welded line pipe prices; second, to develop a valid and representative price prediction model for PT HN; and third, to determine the best method among the three approaches.

2 Literature Overview

The modelling of commodity price forecasting, particularly in the context of pipelines within the oil and gas industry, poses a significant challenge, due to the limited number of studies addressing this topic. In this section, the authors will review several studies related to forecasting similar commodity prices and costs, both in the oil and gas sector and other commodities.

2.1 Previous Research

A comprehensive investigation into the factors influencing the commodity price of pipelines has revealed a multitude of variables that exert a significant impact on pricing. These variables encompass a wide range of economic and market-related factors, including but not limited to oil and gas production, steel prices, transportation costs, and export-import policies among nations. Korneyev's (2022) study demonstrated that a decline in pipeline or petroleum exports can lead to a reduction in pipeline demand and an increase in their production costs [4]. Gruber (2011) further elaborates on this point, underscoring the significance of demand, currency exchange rates, interest rates, inflation, and supply in determining prices [5]. Concurrently, Rui et al. (2011) identified five cost components in pipeline construction: materials, labor, incidental costs, right of way (ROW), and total costs [6]. However, their study centered on a global model that may not be universally applicable to the domestic market.

In regard to the forecasting of pipeline prices, a variety of approaches have been employed in the development of prediction models, yielding disparate results. Amarilies et al. (2022) employed the cost structure method for cost estimation, although this model has not been verified and relies on sensitive company data [7]. Kaiser (2009) developed an empirical model for offshore drilling that has proven to be effective; however, it requires in-depth data for its implementation [8]. Fadjar (2008) combined cost structure with Monte Carlo simulation to handle uncertainty in cost prediction [9]. Concurrently, the implementation of Machine Learning (ML) technology is expanding, as evidenced by the studies conducted by Matel et al. (2022) who employed Artificial Neural Networks (ANN) to calculate the cost of engineering services, and by Xu & Zhang (2023) who utilized Gaussian process regression to predict steel prices [10][11]. Classical approaches, including linear and multivariate regression, have demonstrated efficacy in various studies. For instance, the research by Van Der Steen (2018) and Mahamid (2011) highlighted the superiority of these methods in cost estimation [12][13]. Similarly, Javaid et al. (2022) employed linear regression for flowmeter calibration in the energy sector, underscoring its practical applications [14].

2.2 Time Series Data

Time series data is defined as data that is collected or observed periodically based on a time sequence, such as daily, monthly, or yearly data. In the course of analyzing this data, it is imperative to acknowledge the emergence of patterns. Hanke & Wichen (2005) identified four primary patterns in time series data. The first pattern is the trend, which shows an upward, downward, or steady trend in the long term [15]. The second pattern is seasonality, which is a periodically recurring pattern. The third pattern is cyclical fluctuations influenced by economic or social conditions in a long cycle. The fourth pattern is random or irregular fluctuations, which are fluctuations that cannot be predicted from known factors.

2.3 Exponential Moving Average (EMA)

This approach prioritizes the most recent data points, leveraging exponential averaging to forecast future values. The following mathematical equation (1) is given:

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1} \quad (1)$$

which:
$\hat{y}_{t+1}$ = the forecasting value for the next period,
y_t = observation value in the current period,
$\hat{y}_{t-1}$ = the forecasting value in the previous period,
A = the smoothing factor ($0 < \alpha < 1$) that determines how fast the exponential decay weights.

2.4 Holt's Linear Exponential Smoothing

This method is a development of EMA that is able to handle data with linear trends by adding level and trend components. The following mathematical equation (2), (3), and (4) is given:

$$\hat{y}_{t+h} = l_t + h \cdot b_t \quad (2)$$
$$l_t = \alpha \cdot y_t + (1-\alpha) \cdot (l_{t-1} + b_{t-1}) \quad (3)$$
$$b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1} \quad (4)$$

which:
$\hat{y}_{t+h}$ = the forecast value for h periods ahead,
l_t = component level for period t,
b_t = component trend for period t,
A = smoothing factor for level,
β = smoothing factor for trend.

2.5 Autoregressive Integrated Moving Average (ARIMA)

The ARIMA (or Box-Jenkins) model is a prominent method in time series analysis, comprising three fundamental components: autoregressive (AR), differencing (I), and moving average (MA). This model is employed for data that has been made stationary through a differencing process. The following mathematical equation (5) is given:

$$(1 - \emptyset_1 L - \emptyset_2 L^2 - \cdots - \emptyset_p L^p)(1-L)^d y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\varepsilon_t \quad (5)$$

which:
y_t = the value in the time series at time t,
L = lag operator,
$\emptyset_p$= autoregressive coefficient for lag p,
θ_q = moving average coefficient for lag,
P = AR order,
d = order of differentiation,
q = MA order.

2.6 Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX)

ARIMAX represents an extension of the ARIMA model, incorporating an exogenous variable (X) that exerts an influence on the primary variable (Y) that is being predicted. The following mathematical equation (6) is given:

$$Y_t = \alpha + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=0}^{r} \beta_k X_{t-k} + \varepsilon_t \tag{6}$$

which:

Y_t = the value in the time series at time t,
X_t = oxoge variable (outside factor) at time t,
$\emptyset$ = autoregressive coefficient,
θ = moving average coefficient,
β = coefficient of oxoge variable,
ε_t = error (white noise),
p, d, q = ARIMA parameters.

3 Methodology

3.1 Data Collection

The objective of this research is to obtain historical data on the pricing of welded line pipe from the purchase of API 5L Grade B standard ERW pipes of various sizes (NPS 2"-10") at PT Hulu Nasional Indonesia (PT HN). This data was obtained through the MySAP system for the period January 2019 to February 2025. The selection of this pipe type is predicated on the high and routine frequency of purchases in the company. This research employs a descriptive quantitative approach to construct a pipe price prediction model, in response to the challenges posed by price fluctuations in the market, which complicate tender planning. Furthermore, a comprehensive literature study was conducted to enhance the selection of statistical methods, encompassing Multiple Linear Regression (MLR) and time series analysis, such as ARIMA and ARIMAX.

The data collected encompassed both primary and secondary sources. The primary data set consists of historical pipe prices from MySAP, which have been converted from units per unit joint to units per ton. This conversion was based on the standard weight of ERW pipes Sch 40, with the aim of unifying the scale and increasing the number of samples available. Secondary data encompasses world oil prices, Hot Rolled Coil (HRC) prices, manufacturing labor wages in Indonesia, the rupiah middle exchange rate, and global Welded Pipe Line prices, with reliable sources such as BPS, Bank Indonesia, and Fastmarkets. All variables were aligned to monthly time units for the purpose of modeling, and subsequently interpolated to address the issue of missing data. The adjustment of this time scale and the conversion of units support the accuracy and feasibility of the model in forecasting medium-term prices (three months to three years) according to the guidelines [17].

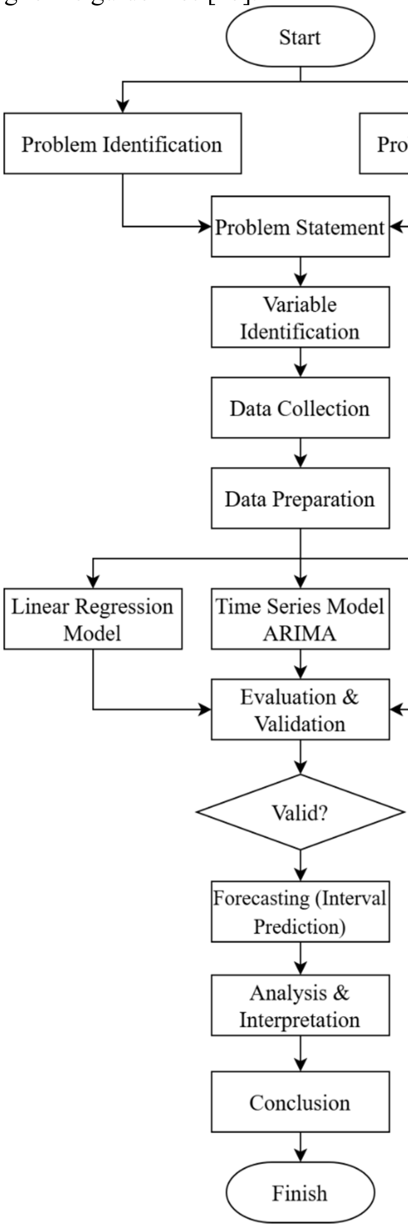

Fig. 1. Methodology Flowchart

3.2 Linear Regression Model Development

The development of a linear regression model is initiated by the examination of the data through the processes of data identification, cleansing, and filtration. Subsequent to this, a series of assumption tests were conducted on each predictor variable, followed by a correlation evaluation between variables to obtain an overview of the relationship between them. Subsequently, all independent variables were incorporated into the model to be evaluated for significance using the partial test (t-test), wherein only significant variables were retained in the final model. Subsequent to this step was a linear regression assumption fulfillment test, which was implemented to ensure the validity of developed model (see Fig. 2).

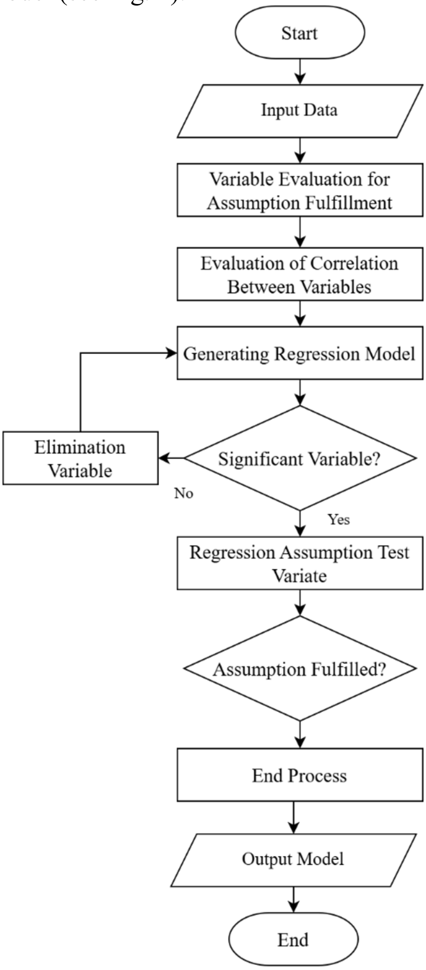

Fig. 2. Linear Regression Model Development

3.3 ARIMA Model Development

In addition to causal modeling, time series data analysis is also carried out using ARIMA (Auto Regressive Integrated Moving Average), a step by step process that is carried out is described in Fig. 3.

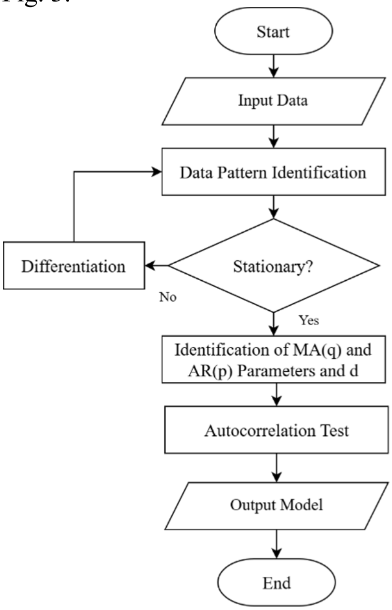

Fig. 3. ARIMA Model Development

3.4 ARIMAX Model Development

ARIMAX is a forecasting model that utilizes external variables to predict time series data. These external variables are hypothesized to influence the primary variable that is being predicted. This model represents an extension of the ARIMA model, which relies exclusively on patterns in historical data, by incorporating exogenous (outside) variables to enhance prediction accuracy.

3.5 Model Evaluation and Validation

The evaluation of the model is achieved through a comparison of the model's predicted results with the actual data, followed by the calculation of the error. The performance measures employed include Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Percentage Error (MAPE), and R2. In order to assess the overall performance of the model, cross-validation techniques are employed. The validation of the regression model was conducted by taking into account the time lag (t-1) in order to predict the value at time t.

4 Results and Discussion

4.1 Identification of Dependent and Independent Variables

The dependent variable in this study is the price of welded line pipe at PT HN, while the independent variables consist of the price of raw materials, labor wages, dollar exchange rates, crude oil prices, and material delivery locations. The dependent variable is defined as the variable that is measured in response to changes in the independent variable [16]. In accordance with the findings of preceding studies, including those undertaken by Inna (2006), the factors influencing the cost of steel products in Ukraine encompass the costs of raw materials, labor wages, manufacturing capacity, and the exchange rate of the U.S. dollar [19]. Concurrently, Hendro Asmoro (2017) determined that raw material and labor costs comprise the most substantial components in determining the price of welded line pipe [20]. Furthermore, the House of Risk approach by Arwan Kholid & Harito (2024) demonstrated that crude oil price fluctuations constituted the primary source of risk in pipe procurement at Pertamina EP [21].

The results of the global market trend analysis in Table 1 also demonstrate a strong correlation between global welded line pipe prices and raw material (HRC) prices, with an R^2 value of 0.893, indicating a significant influence. In light of these findings, the researcher also considered the input from Ivan (2022), who underscored the significance of transportation factors and regional policies in the determination of pipe product prices [22]. Consequently, this study incorporates the delivery location variable as an indicator of logistics costs, given the discrepancy in distribution costs to Sumatra, Java, and Eastern Indonesia (KTI). The extent of the delivery location's remoteness directly correlates with the incurred logistics costs, which may result in an increase in the final cost of the pipe.

Table 1. Linear Regression Between Global Pipe Prices and Raw Material Prices

Model Summary				
R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
945a	0.893	0.892	42.14996	0.650
a. Predictors: (Constant), HRC (USD /Ton)				
b. Dependent Variable: Welded Line Pipe Global (USD /Ton)				

4.2 Data Preparation

4.2.1 Conversion of Dependent Variable

The price of the pipe was converted from per joint to per ton to simplify and generalize the model, thus enabling its application to a wide range of pipe sizes (NPS) and lengths (SLR/DLR). This conversion is consistent with the ASME standard, which provides a quantitative listing of pipe weight in kilograms per meter for each specification, as illustrated in Table 2.

Table 2. Conversion Based on ASME B36.10M-2004

No	Nominal Pipe Size (NPS) Sch 40	Kg / m ref: ASME (2004)	Ton / Joint (1 join = 6 m)
1	2"	5.44	0.03
2	3"	11.29	0.07
3	4"	16.08	0.10
4	6"	28.26	0.17
5	8"	42.55	0.26
6	10"	60.29	0.36

Table 3 and Table 4 present the descriptive statistics before and after unit conversion, respectively. The SPSS output results demonstrate that the pipe price per ton, subsequent to conversion, exhibits relative uniformity across the various pipe sizes. To ascertain whether the converted data originated from a population with the same mean, a comparison test of mean and standard deviation was conducted using the One Way ANOVA method.

Table 3. Descriptive Statistics Before Unit Conversion

Descriptive Statistics (Pipe Price per Join, 1 Join = 6m)						
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
2in	47	51.74	77.03	63.1177	8.80828	77.586
3in	103	109.85	318.31	190.5672	68.88280	4744.841
4in	71	177.21	445.86	280.2539	96.35988	9285.227
6in	47	270.81	438.66	344.3685	41.19737	1697.224
8in	17	426.55	660.44	527.8182	83.09898	6905.440
10in	6	594.07	970.42	841.1250	140.42982	19720.535

Table 4. Descriptive Statistics After Unit Conversion

Descriptive Statistics (Pipe Price per Ton)						
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
2in	47	1585.28	2360.09	1933.7723	269.87049	72830.081
3in	103	1583.50	2367.93	1929.9557	232.44223	54029.390
4in	71	1624.34	2588.04	2003.2152	251.86476	63435.859
6in	47	1597.12	2587.02	2030.9553	242.96666	59032.799
8in	17	1670.78	2586.93	2067.4441	325.49741	105948.564

However, it is imperative to conduct a normality test prior to conducting an analysis of variance (ANOVA). The results of the Shapiro-Wilk test indicated that the majority of the initial data were not normally distributed. Consequently, a transformation was executed using the Box-Cox method. Subsequent to the transformation, all subsets satisfied the normality assumption ($\alpha > 0.05$), thereby enabling further analysis of the data using One Way ANOVA. The ensuing results were obtained from the aforementioned test as illustrated in Table 5.

Table 5. One Way ANOVA Test

ANOVA (TrfPriceTon)					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	573557.013	4	143389.253	2.412	.05
Within Groups	16523798.481	278	59438.124		
Total	17097355.494	282			

The One Way ANOVA test results on the pipe price per ton data show a significance value (α) of 0.05, which means H_0 is accepted, indicating that there is no significant difference between the pipe diameter groups. Subsequent to this initial analysis, a post-hoc test employing the Tukey HSD method was utilized to ascertain specific disparities among pipe diameter groups. In the event that the Tukey results demonstrate that all groups are within the same average group, it can be concluded that the diameter variation does not significantly affect the pipe price per ton.

Table 6. Turkey HSD Test

Tukey HSDa,b (TrfPriceTon)		
Label	N	Subset for alpha = 0.05
		1
3in	103	1935.8790
2in	47	1945.1640
4in	71	2012.5804
6in	46	2030.8350
8in	16	2067.2638
Sig.		.127

As demonstrated in Table 6, the Turkey HSD test's output reveals that all categories are consolidated into a single average group. Therefore, the variation in the size of the inner diameter of the welded line pipe is validly aggregated into a single variable: the PT HN pipe price per ton.

4.3 Linear Regression Causal Model Development

The data set under consideration contains 334 items. The delivery location variable (Sumatra, Java, KTI) is a nominal variable that is converted into a dummy variable (0/1), with Java serving as the reference, as illustrated in Table 7.

Table 7. Variable Descriptive Statistics

Descriptive Statistics				
Variable		Mean	Std. Deviation	N
Price/Ton	X_1	2,006.43	258.74	334
Kurs USD	X_2	14,804.70	701.51	334
Brent Oil	X_3	74.09	19.16	334
Upah	X_4	4,743,784.56	317,506.09	334
HRC	X_5	602.18	136.06	334
KTI	X_{6_1}	0.30	0.46	334

Table 8 shows the SPSS output for the multiple linear regression analysis model with the dependent variable X_1 and independent variables X_2, X_3, X_4, X_5, dan X_{6_1}, X_{6_2}.

Table 8. Regression Linear Model with 5 variables

	Unstdzd Coefficients	Stdz Coef-ficients	t	Sig.	Collinearity Statistics	
	B	Beta			Tolerance	VIF
Const	-1351.570		-4.70	0.00		
X_2	0.205	0.557	15.62	0.00	0.705	1.419
X_3	7.131	0.528	13.17	0.00	0.557	1.794
X_4	-3.75E-05					
X_5	-0.046	-0.83	0.41	0.289	3.464	
X_{6_1}	-0.167	-0.088	-1.46	0.14	0.248	4.028
X_{6_2}	183.102	0.324	7.15	0.00	0.438	2.285

R	R^2	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
0.841	0.707	0.702	141.28870	0.980

4.3.1 Partial Significance Test (t test)

Pursuant to the implementation of the significance test, it was ascertained that the value of variables X_4, X_5, X_{6_2} exceeds 0.05 (Table 8). Consequently, the null hypothesis (H0) is thereby accepted. This finding indicates that X_4, X_5, X_{6_2} variable exerts negligible influence on the model's predictions. This is evidenced by the minimal change in the R^2 value following the elimination of the variables in the construction of the linear regression model. Tabel 9 (SPSS output) below presents the results after the exclusion of X_4, X_5, and X_{6_2} variables.

Table 9. Regressiom Linear Model with 3 Variables

	Unstdzd Coefficients		Stdz Coef-ficients	t	Sig.	Collinearity Statistics	
	B	Std.Error	Beta			Tolerance	VIF
(Constant)	-1641.321	167.395		-9.805	0.000		
x_2	0.210	0.012	0.568	17.992	0.000	0.901	1.110
x_3	6.689	0.434	0.495	15.406	0.000	0.870	1.150
$x_{6.1}$	165.032	17.545	0.292	9.406	0.000	0.935	1.070
(Constant)	-1641.321	167.395		-9.805	0.000		

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
.839[a]	0.703	0.701	141.5768	0.968

4.3.2 Simultaneous Significance Test (F-Test)

The F-test's significance level is derived from the corresponding value listed in Table 10. When the p-value is less than 0.05, it can be concluded that the independent variable has a simultaneous effect on the dependent variable.

Table 10. F-Test

ANOVA (TrfPriceTon)					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	15678030.07	3	5226010.02	260.727	.000[b]
Residual	6614516.48	330	20043.99		
Total	22292546.55	333			

4.3.3 Heterodasticity Test

The data points are scattered above and below, or around, the value of 0, and no specific pattern is observed in their distribution. Therefore, it can be concluded that there is no heteroskedasticity problem, as shown in Fig. 4.

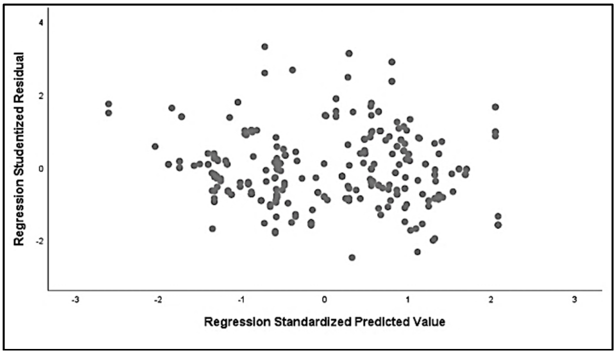

Fig. 4. Heteroscedasticity Test

4.3.4 Multicollinearity Test

The multicollinearity test is indicated by the VIF values. As shown in Table 9, the VIF values for the three independent variables are less than 10, indicating that multicollinearity is not an issue.

4.3.5 Normality Test

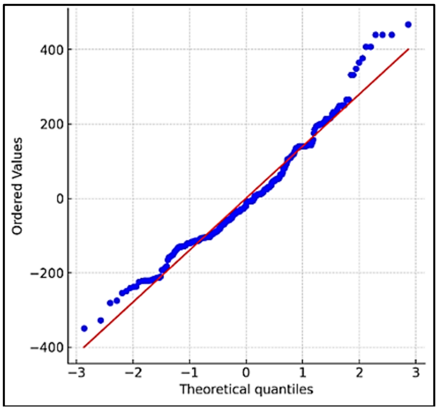

Fig. 5. Normality Test

As illustrated in Figure 5 above, The Shapiro-Wilk normality test results show a p-value of 1.39×10^{-6} which is smaller than 0.05. This indicates that the residuals are not normally distributed To address this issue, the dependent variable (X1) was transformed using the Box-Cox transformation (with $\lambda = -2$).

4.3.6 Autocorrelation Test

The condition for no autocorrelation is if the Durbin-Watson (d) value of the regression model falls between the dU and (4-dU) values. Based on Table 9, with (k'; N) = (3;

334), the values of dL = 1.84 and dU = 1.88 are obtained, while the Durbin-Watson (d) value of the regression model is 0.968, which does not meet the requirement, indicating autocorrelation in the regression model. As the remedy, the data was transformed by shuffling the dataset order, so that the DW value became 1.847 as shown in Table 11. Thus, the Autocorrelation Test has been met, indicating no autocorrelation in the regression variables.

Preliminary regression analysis indicates that the price of pipes is influenced by three variables: the price of Brent Oil, USD/IDR exchange rate, and the delivery location within the KTI Region. The geographical location of material delivery to KTI has been demonstrated to exert an influence on pipe prices, while Java and Sumatra have been found to be unaffected by this phenomenon. Multiple linear regression model can be formulated with constant and coefficients regression as shown in Table 11.
which:
X_1 = pipe price ERW/ton
X_2 = USD to IDR exchange rate
X_3 = oil price (Brent Oil)
X_{6_1} = 0 for other than KTI, 1 for KTI.

Table 11. Coefficients Regression after Transformation

	Unstdzd Coefficients		Stdz Coef-ficients	t	Sig.	Collinearity Statistics	
	B	Std.Error	Beta			Tolerance	VIF
Constant	0.500	0.000		0.000	Constant	0.500	0.000
X_2	2.628E-11	1.434E-12	0.560	2.824E-52	X_2	2.628E-11	1.434E-12
X_3	8.916E-10	5.392E-11	0.515	3.373E-45	X_3	8.916E-10	5.392E-11
X_{6_2}	2.167E-08	2.169E-09	0.300	1.020E-20	X_{6_2}	2.167E-08	2.169E-09
Constant	0.500	0.000		0.000	Constant	0.500	0.000
Dependent Variable: Box-Cox (Pipe/ton,-2)							

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
.838	0.702	0.70	141.87310	1.847

4.4 Regression Model Evaluation

In order to assess the reliability of the model, a process of validation was executed by employing cross-validation techniques, which entailed the comparison of 80 % of the training data and 20 % of the test data. The validation results indicated a maximum attainable prediction error (MAPE) of 5.4 % for the training set and 5.84 % for the test set. The low and stable MAPE value suggests that the regression model with three variables is valid. The validity of the model must be maintained through regular validation procedures.

4.5 Time Series Model Development ARIMA Method

In addition to employing the causal model approach, this study also utilized the Autoregressive Integrated Moving Average (ARIMA) model to analyze time series data.

4.5.1 Identification of Stationary Data Patterns

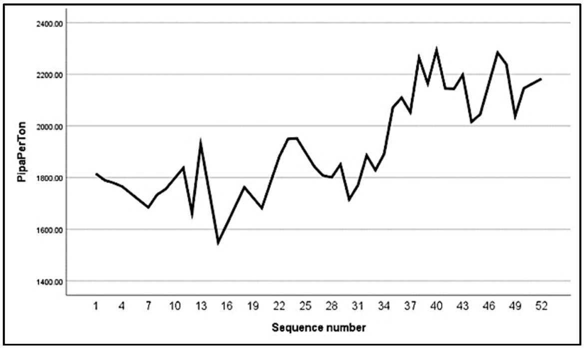

Fig. 6. Sequence Diagram Pipe Price/Ton

As illustrated in Figure 6, the data pattern is not stationary, indicating the necessity of performing differentiation 1. As illustrated in Figure 7, the data plot following the initial differentiation process appears to be stationary, a conclusion that can be drawn based on a visual assessment of the data. As illustrated in Figure 8 and Figure 9, the autocorrelation function (AFC) plot reveals a stationary data pattern when the AFC value falls within the confidence interval limits.

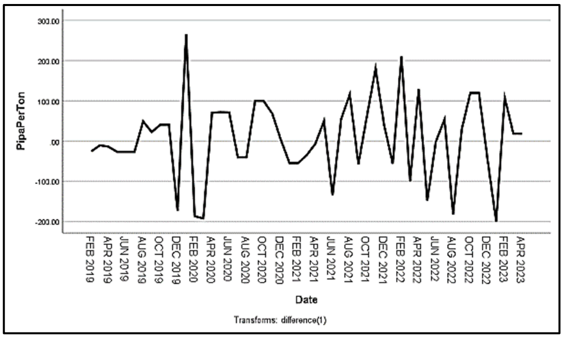

Fig. 7. Sequence Diagram Pipe Price/Ton (Differentiation)

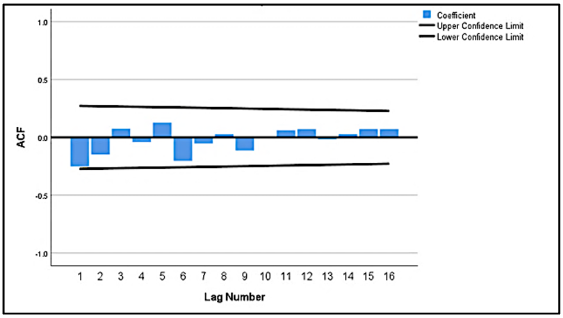

Fig. 8. AFC Graph

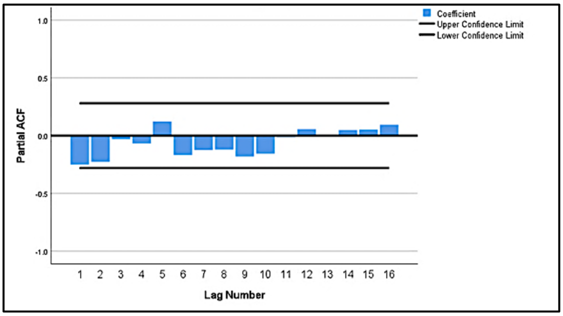

Fig. 9. Partial AFC Graph

4.5.2 Determination of MA(q) and AR (p) parameters

Table 12. Comparison of ARIMA (p,d,q) Combinations

	A : ARIMA (1,1,1)	B : ARIMA (0,1,1)	C :ARIMA (2,1,2)
Norm. BIC	9.56	9.50	9.68
MAPE	3.68	3.87	3.70
RMSE	108.96	109.02	108.96
MAE	74.52	78.21	74.51
R^2	0.80	0.79	0.80

The subsequent step involves the determination of MA(q) and AR(p) orders, which are derived from the ACF and PACF plots. A comparative analysis is performed on several combinations of ARIMA (p, d, q) models using various evaluation metrics, such as normalized BIC, MAPE, RMSE, MAE, and R^2. A comparison of the evaluation results led to the selection of the ARIMA (1,1,1) model as the optimal model. Despite exhibiting a lower BIC value in comparison to the ARIMA (0,1,1) model, the ARIMA (1,1,1) model demonstrates the most optimal performance in other evaluation metrics, including MAPE, RMSE, MAE, and R^2 as explained from Table 12. This configuration is optimal in terms of achieving a balance between accuracy and efficiency. The ARIMA statistical model is delineated in Table 13 and Fig. 10.

Table 13. ARIMA Statistical Model

Model Statistics					
Model	Number of Predictors	Ljung-Box Q(18)			Number of Outliers
		Statistics	DF	Sig.	
PipaPerTon-Model_1	0	13.752	16	0.617	0

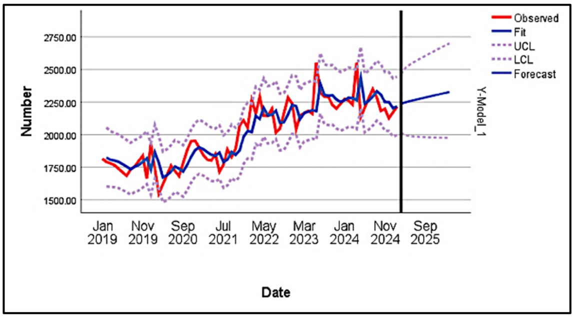

Fig. 10. ARIMA Model

4.5.3 Autocorrelation Test

To ensure that there is no autocorrelation in the residuals, the Ljung-Box test was conducted. The results indicate a p-value of 0.617 (> 0.05), suggesting an absence of substantial evidence supporting the hypothesis of autocorrelation. Consequently, the residuals of the ARIMA (1,1,1) model can be regarded as white noise, thereby satisfying the fundamental assumptions of the time series model.

4.5.4 Validation of ARIMA Model

The validation of the ARIMA (1,1,1) model was conducted using a cross-validation technique, with 80 % of the data utilized for training and the remaining 20 % allocated for testing purposes. Consequently, the mean absolute percentage error (MAPE) of the training set is 3.68 %, and the test set is 5.64 %, indicating a relatively high degree of accuracy and stability. Consequently, the model is regarded as valid; however, its validation is required to be conducted on a regular basis to ensure its reliability.

4.6 Time Series and Causal Model Development (ARIMAX)

4.6.1 Analysis of Exogenous Variables in The ARIMAX Model

The ARIMAX analysis was conducted by combining the time series model and the exogenous variables suspected of causally affecting Welded Line Pipe prices, namely the USD/IDR exchange rate and Brent oil price. The delivery location variable was not included because it did not conform to the monthly aggregation format employed in the time series analysis. The objective of this analysis is to ascertain whether the incorporation of external variables can enhance the performance of the previously constructed ARIMA model. The ARIMAX statistical model is delineated in Table 14 and Fig. 11.

Table 14. ARIMAX Statistical Model

Model Statistics								
Number of Predictors	Model Fit Statictics					Ljung-Box Q(18)		
	R^2	RMSE	MAPE	MAE	Normalized BIC	Statistics	DF	Sig.
2	0.815	105.765	3.530	20.785	9.616	16	0.616	0.618

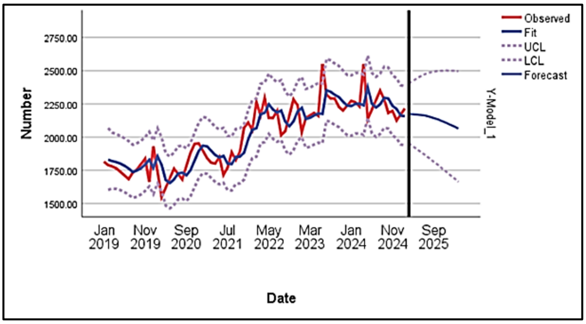

Fig. 11. ARIMAX Model

4.6.2 Validation of ARIMAX Model

The validation of the ARIMAX (1,1,1) model is accomplished through cross-validation, employing 80 % of the training data and 20 % of the test data. The results indicate a mean absolute percentage error (MAPE) of 3.68 % on the training set and 3.00 % on the test set. This value is relatively low and stable, which leads to the model being declared valid. Nevertheless, the necessity of periodic validation persists to ensure the system's reliability over time.

4.7 Evaluation of Regression, ARIMA and ARIMAX Models

The present study compares three models: multiple linear regression, ARIMA(1,1,1), and ARIMAX(1,1,1) with two exogenous variables (the USD exchange rate and the Brent oil price). Subsequent to undergoing a series of assumption and validation tests, their performance is evaluated based on the test data. The findings indicate that the ARIMAX (1,1,1) model demonstrates superior performance, exhibiting the lowest values for MAPE, RMSE, and MAE as shown in Table 15. Regression models remain a valuable tool for interpreting the relationship between variables, while ARIMA can be employed when exogenous variables are not available.

The following visual representation in Fig. 12 and Fig. 13, illustrates the comparison between the actual data and the prediction results of the three models.

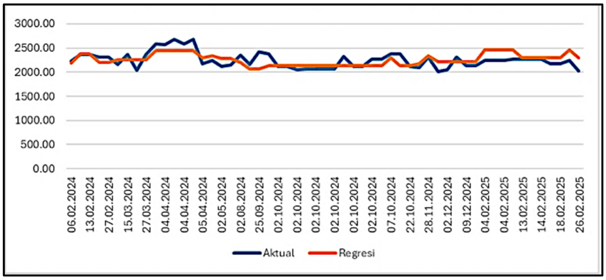

Fig. 12. MAE, RMSE, and MAPE Comparison with Regression

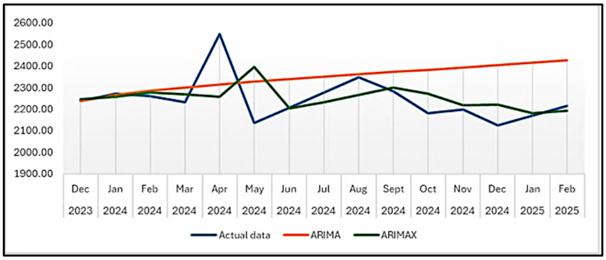

Fig. 13. *MAE, RMSE, and MAPE Comparison with ARIMA, ARIMAX*

Table 15. MAE, RMSE, and MAPE Comparison

Model	Method	Predictor	MAPE	RMSE	MAE
1	Multiple Linear Regression	3 variable	5.84 %	153.6941	130.6866
2	ARIMA (1,1,1)	-	5.87 %	160.3875	130.5303
3	ARIMAX (1,1,1)	2 variable	2.95 %	109.5012	67.35505

4.8 Forecasting Results

As illustrated in Table 16 and Fig. 14, the results of the Welded Line Pipe price prediction interval at PT HN for the period March 2025 to March 2026 were determined using the ARIMAX (1,1,1) model with two primary predictors: the world crude oil price (Brent oil) and the USD/IDR exchange rate. The LCL (lower confidence limit) denotes the lower limit, while the UCL (upper confidence limit) signifies the upper limit of the estimated price. This establishes a price range that can be utilized as a reference point for the Owner Estimate in assessing the reasonableness of the pipe purchase price by PT HN.

Preliminary analyses of available forecasting data indicate a downward trend in pipeline prices over the ensuing 12-month period. This finding aligns with the observed downward trend in global oil prices, a key exogenous variable incorporated into the model. However, the rate of decline in pipe prices is not precipitous due to its continued influence by the elevated USD/IDR exchange rate, wherein the depreciating rupiah tends to impede price reductions in the context of import-based acquisitions or global commodities.

Table 16. Forecasting Result

Year	Month	Pipe Price Prediction (USD / Ton)			Price Index (reference : 1/2025)
		Prediction	LCL	UCL	
2025	1	2180.73	1970.82	2398.49	1
	2	2192.76	1982.28	2411.09	1.01
	3	2205.35	1994.28	2424.28	1.01
	4	2200.42	1969.02	2441.32	1.01
	5	2198.21	1962.63	2443.66	1.01
	6	2192.35	1956.04	2438.6	1.01
	7	2186.56	1950.28	2432.81	1.00
	8	2175.29	1939.54	2421.02	1.00
	9	2161.67	1926.65	2406.68	0.99
	10	2149.04	1914.71	2393.36	0.99
	11	2130.77	1897.46	2374.07	0.98
	12	2114.24	1881.86	2356.61	0.97
2026	1	2093.15	1861.97	2334.32	0.96
	2	2074.67	1844.54	2314.79	0.95
	3	2052.99	1824.1	2291.87	0.94

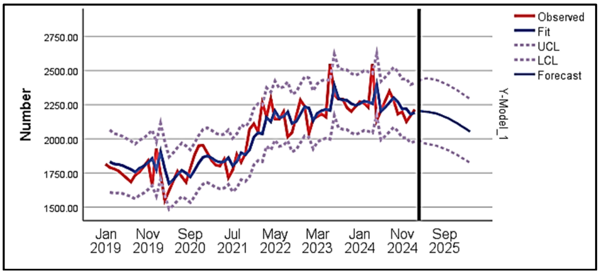

Fig. 14. Forecasting Result

5 Conclusion

A thorough examination of the available data, employing rigorous linear regression analysis, reveals a substantial positive correlation between the prices of global welded line pipe and HRC raw materials, with an R^2 value of 0.9. However, this relationship does not apply to the price of Welded Line Pipe at PT HN, because the t-test results show that the price of HRC raw materials has no significant effect. In contrast, the multiple linear regression results show that the price of Brent Oil, the dollar exchange rate, and the location of shipments to East Kalimantan Indonesia (KTI) have a significant positive correlation to the price at PT HN, with an adjusted R^2 value of 0.7. The most significant influence is attributed to the dollar exchange rate (0.568), followed by crude oil prices (0.495), and delivery location (0.292).

PT HN's price discrepancy on the global market is attributable to the distinctive characteristics of Indonesia's oil and gas industry, which is characterized by elevated barriers to entry. Factors such as the obligation to fulfil TKDN, strict HSSE standards, and the need for large volumes of materials restrict entry to only certain PBJs in this market. To enhance the precision of price forecasting, three approaches were employed: Multiple Linear Regression, ARIMA, and ARIMAX. The optimal outcomes were attained with the ARIMAX (1,1,1) model, incorporating exogenous variables of the USD/IDR exchange rate and Brent Oil price, yielding a MAPE of 2.95 %, which is deemed sufficient for OE price validation at PT HN.

References

1. SKK Migas. (2023a). Juklak Pengadaan Barang Jasa PTK007 Buku Kedua Rev 05.
2. SKK Migas. (2023b). Pedoman Tata Kerja Nomor: PTK-007/SKKIA0000/2023/S9 (Revisi 05).
3. Hendro Asmoro, T. (2017). Analisis Strategi Pengembangan Industri Pipa Baja di Indonesia.
4. Korneyev, M. (2022). Economic Factors Influencing Pipeline Market in Eastern Europe.
5. Gruber, H. (2011). Applied Industrial Economics.
6. Rui, Z., Chen, J., & Wang, Y. (2011). Pipeline Construction Cost Analysis: A Global Perspective.

7. Amarilies, D., Nugraha, A., & Wijaya, R. (2022). Cost Structure Approach for Oil and Gas Projects.
8. Kaiser, M. J. (2009). Offshore Cost Estimation Model for Drilling and Production.
9. Fadjar, A. (2008). Implementasi Simulasi Monte Carlo untuk Estimasi Biaya Proyek.
10. Matel, A., Siregar, F., & Yusuf, A. (2022). ANN Application in Engineering Cost Prediction.
11. Xu, H., & Zhang, Y. (2023). Steel Price Forecasting Using Gaussian Process Regression.
12. Van Der Steen, R. (2018). Multivariate Regression Models for Infrastructure Costing.
13. Mahamid, I. (2011). Common Causes of Cost Overrun in Construction.
14. Javaid, M., Khan, I. A., & Ahmed, W. (2022). Flowmeter Calibration Using Linear Regression in Energy Sector.
15. Hanke, J. E., & Wichern, D. W. (2005). Business Forecasting (9th ed.).
16. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate Data Analysis (7th ed.).
17. Heizer, J., & Render, B. (2014). Operations Management (11th ed.).
18. Inna, L. (2006). Cost Analysis of Steel Production in Ukraine.
19. Hendro Asmoro. (2017). Analisis Struktur Biaya Produksi Pipa Line Welded di Indonesia.
20. Arwan Kholid, & Harito. (2024). Mitigasi Risiko Pengadaan Pipa di Pertamina EP Menggunakan House of Risk.
21. Ivan, R. (2022). Determinants of Steel Product Pricing in Indonesia.