A MIXED DATA SAMPLING (MIDAS) APPROACH TO MODELING CRUDE OIL PRODUCTION IN NIGERIA: ACCOUNTING FOR MIXED—FREQUENCY DATA

Amalahu, Christian Chinenye[1] and Ekeadinotu, Chibueze Barnabas[2]

[1]Department of Mathematics, University of Agriculture and Environmental Sciences, Umuagwo Imo State, Nigeria.
E-mail: chinenyechris2017@gmail.com

[2]Department of Mathematics, University of Agriculture and Environmental Sciences, Umuagwo Imo State, Nigeria.
E-mail: chibuezeekeadinotu@gmail.com

ABSTRACT

This research work attempt to establish an efficient method of forecasting Nigeria' Crude Oil Production and the Nigeria Gross Domestic Product applying Mixed Data Sampling approach (MIDAS) from 2010 to 2022. It combined data set of different frequencies; quarterly and monthly in the same regression. It was observed that based on the Root Mean Square Error the MIDAS Almon (PDL) regression model provided a better model estimation than the MIDAS Step weighting and the MIDAS Beta. The monthly Crude oil production has a positive effect on GDP as the slope coefficient is statistically significant 0.849895 (Prob. 0.0000). It is therefore important for appropriate policy formulation and implementation of such policies to encourage and boost these variables for effective management of Crude oil production in Nigeria. Hence, direct relationship between Crude oil production and GDP is needed to diversify the economy base to enhance productive activities in Nigeria and better crude oil production.

Key words: Mixed Data Sampling, MIDAS Almon (PDL), MIDAS Beta

1. INTRODUCTION

Time Series Regression traditionally use data where all variables are sampled at the same frequency while Mixed Data Sampling (MIDAS) Regression accommodates data sampled at varying frequencies to be adopted in the same regression. In mixed frequencies settings, temporal aggregation is commonly applied to equate the variables; in the process a lot of information might get loss when higher-frequency variables are transformed to lower-frequency variables. This raises a need of models that can incorporate and take use of the potentially large information from high-frequency data. To address this problem, Ghysels *et al.*, (2004) introduce Mixed Data Sampling (MIDAS) regression models which allow for the regressand and the regressors to be of different frequencies. Their primary goal was to widen the volatility literature, Ghysels *et al*., (2006), the model's versatility has been remarkably broad, comprising modeling of macroeconomic aggregates, such as inflation, GDP and rate of unemployment. It is designed to find a balance between retaining the individual timing information of the high frequency data and reducing the number of parameters that need to be estimated. It is believed to have better estimating and forecasting ability than many other conventional models Tsui *et al*., (2013).

In a MIDAS regression, a large number of lagged values of the high-frequency variables can be considered without restrictions. On the lagged polynomial, the number of parameters to estimate

ISSN NUMBER: 1116-249X

might become large. A parsimonious model is found by capturing the coefficients with a known function.

Crude oil is one of the natural resources to mankind and also a major source of energy in Nigeria. It is an important commodity in the world market, not minding the campaign for a green energy and other sources of power. Crude oil is one of the most expensive commodities in the international market and the Nigerian economy is heavily dependent on it Acha *et al*., (2023). Nigeria at this turning point has to harness the potential of its crude oil production to drive economic resilience while grabbing sustainable practices in order to mitigate climate change.

Crude oil production forecasting in Nigeria is crucial for energy planning, policy-making, and decision-making in the oil industry. However, crude oil production data is often available at different frequencies (e.g., monthly, quarterly, annually), which can make it challenging to model and forecast accurately. Traditional models may not fully explore the complex relationships between variables or account for the mixed-frequency nature of the data. This study aims to address this problem by exploring the application of Mixed Data Sampling (MIDAS) models to crude oil production forecasting.

2. LITERATURE REVIEW

Recent studies have showed the effectiveness of MIDAS models in crude oil market analysis. Lyu *et al*. (2024) combined GARCH-MIDAS models with extreme value theory (EVT) to forecast the Value-at-Risk (VaR) of the crude oil market, achieving superior performance compared to benchmark models. They showed that GARCH-MIDAS models outperform benchmark models in forecasting crude oil market volatility, particularly when incorporating suitable low-frequency macroeconomic variables and also revealed that Extreme Value Theory (EVT) improves VaR forecasting accuracy when combined with GARCH-MIDAS models.

Audrone *et al*., (2023) utilized standard volatility models such as Generalized Autoregressive Conditional Heteroskedastic (GARCH), Generalized Autoregressive Score (GAS), and Stochastic Volatility (SV), along with Mixed Data Sampling (MIDAS) regressions, which incorporate the impacts of relevant financial/macroeconomic news into asset price movements. They employed an innovative Bayesian estimation approach called the density-tempered sequential Monte Carlo method. Their findings indicated that the inclusion of exogenous variables was beneficial for GARCH-type models while offering only a marginal improvement for GAS and SV-type models. They also discovered that GAS-family models exhibit superior performance in terms of in-sample fit, out-of-sample forecast accuracy, and Value-at-Risk and Expected Shortfall prediction.

Hassan and Ismail (2023) proposed a quantile regression neural network combined with unrestricted MIDAS, showcasing improved forecasting accuracy. They adopted hybrid QRNN-U-MIDAS model to forecast quarterly GDP applying monthly and weekly data. Their study

ISSN NUMBER: 1116-249X

revealed that Quantile Regression Neural Network enhances forecasting accuracy when combined with unrestricted MIDAS.
Bonnier (2022) forecast crude oil volatility modeling and forecasting. He comparrd multiple GARCH-family models (GJR, EGARCH, log-GARCH, log-GARCH-X) for crude oil volatility forecasts (1, 2, 21, 63 days) and he found out that OVX (CBOE Crude Oil Volatility Index) was a better predictor, with assorted exchange rates staying in the final model and when neither supply/demand factors nor their volatilities were statistically significant.

Tsui *et al*., (2013) investigated the forecasting performance of the three models, i.e., the MIDAS regression model, the direct regression model on high frequency data and the time-averaging regression model, by using data from the Singapore economy. Their results showed that MIDAS regression using high frequency stock returns data produces better forecast of GDP growth rate than the other models, and the best forecasting performance was achieved using weekly stock returns. It was found that the intra-period MIDAS model outperforms other forecasting models, as it can capture well all the important up-and-downs of the economic performance in Singapore, especially during the economic crises in 2001-2002 and 2008-2009, and the best forecasting performance was achieved by using weekly stock returns.

Adeniji *et al*., (2017) investigated the long and short run relationships between broad money supply and real aggregate output (GDP) in Nigeria from 1981 to 2015. They investigated the mystification whether or not money supply as the major monetary policy really impact on the Nigerian economy. Their research work made use of data set of different frequencies (yearly and quarterly) in order to unfold some unknown facts that data set of the same frequency may fail to present. They also employed unrestricted Mixed Data Sampling (U-MIDAS) technique and Autoregressive Distributed Lag (ARDL) technique. The ADF unit root test revealed that the yearly real GDP and quarterly broad money supply contained a unit root and this allow the testing of cointegration among the variables. The U-MIDAS results confirmed the existence of a long and short run relationship between yearly real GDP and quarterly broad money supply at different season, and ARDL result show that money supply impacted significantly on real GDP in the long run. Their study concluded that the disequilibrium correction terms from the two approaches showed the evidence that there is a tendency for growth targeting in Nigeria which is one of the major aim of Nigeria economy though at a very slower rate. They recommended that monetary authority should maintain the level of inflation targeting in the economy and the volume of money to be supplied should be monitored as too much money supply in the economy will lead to hiper inflation and also the periodic money multiplier should be made efficient by supplying the money into the circulation regularly so as to co-trend with the real GDP growth by making cash available for business transactions and other economic activities, this will improve the real GDP of Nigeria economy.

Hence, virtually all the studies on Nigeria crude oil production forecasting involve the use of uniform data frequency. Nonetheless, to the best of our knowledge, no study has applied the

ISSN NUMBER: 1116-249X

MIDAS approach to account for the role of higher frequency GDP in crude oil production forecasting. This study aims to develop a MIDAS model that addresses these challenges and evaluates its performance in forecasting crude oil production.

3. METHODOLOGY

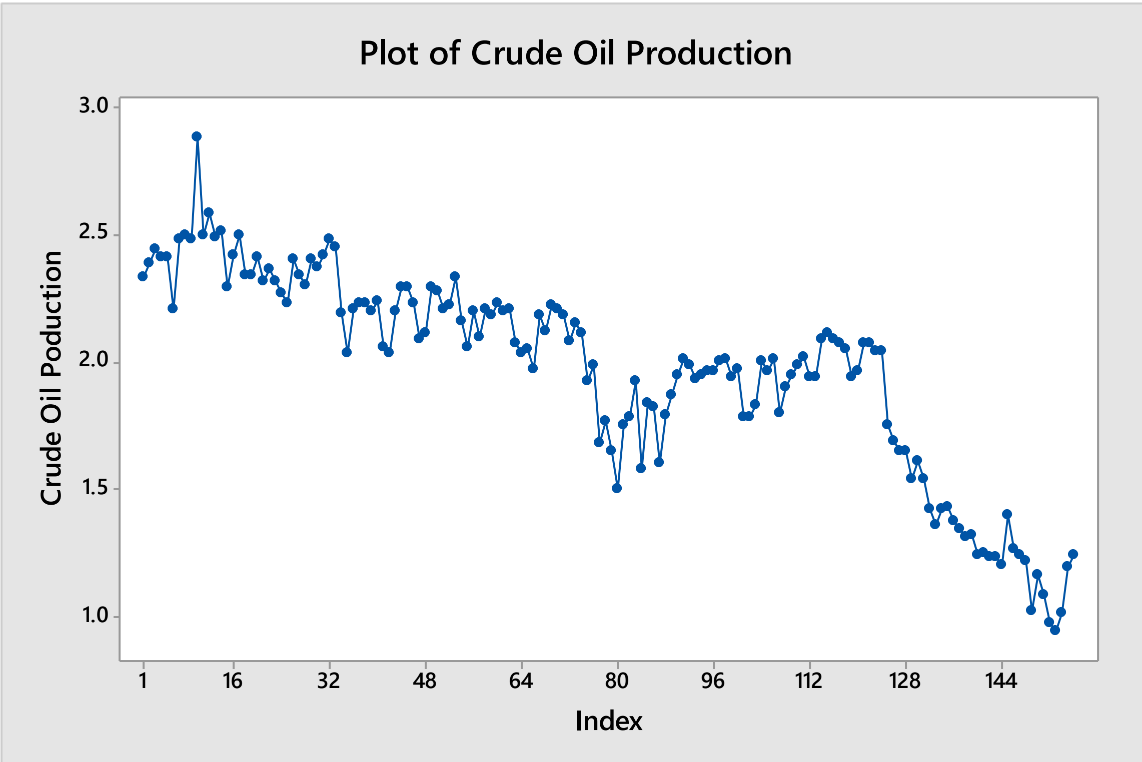
In this research work, we applied Nigerian quarterly GDP data and monthly Nigerian Crude Oil Production data. The relationship between these two variables is that Crude Oil Production is a solution to economic growth and GDP contributes immensely to the Crude Oil Production position in Nigeria. It is expected that an increase in a country's income will also result in an increase of its GDP. We use the Mixed Data Sampling (MIDAS) regression model to deal with a period or frequency difference issues of GDP and Crude Oil Production variables.

The Beta weight function and the exponential Almon weight function are used to estimate the parameters in the mixed data sampling regression model. Their results are compared with the lag distributed model and the model that gives the smallest error is then selected to estimating the GDP of Nigeria.

3.1 The Data and Model

The data used for this research work cover quarterly Gross Domestic Product (GDP) and monthly Nigerian Crude Oil Production for the period spanning Q1 2010 to Q4 2022.
The log-transformed series is subdivided for estimation and forecast purposes.
The data for the study were sourced from the various issues of the statistical Bulletin of Central Bank of Nigeria (CBN). It is represented graphically in Fig. 1.

ISSN NUMBER: 1116-249X



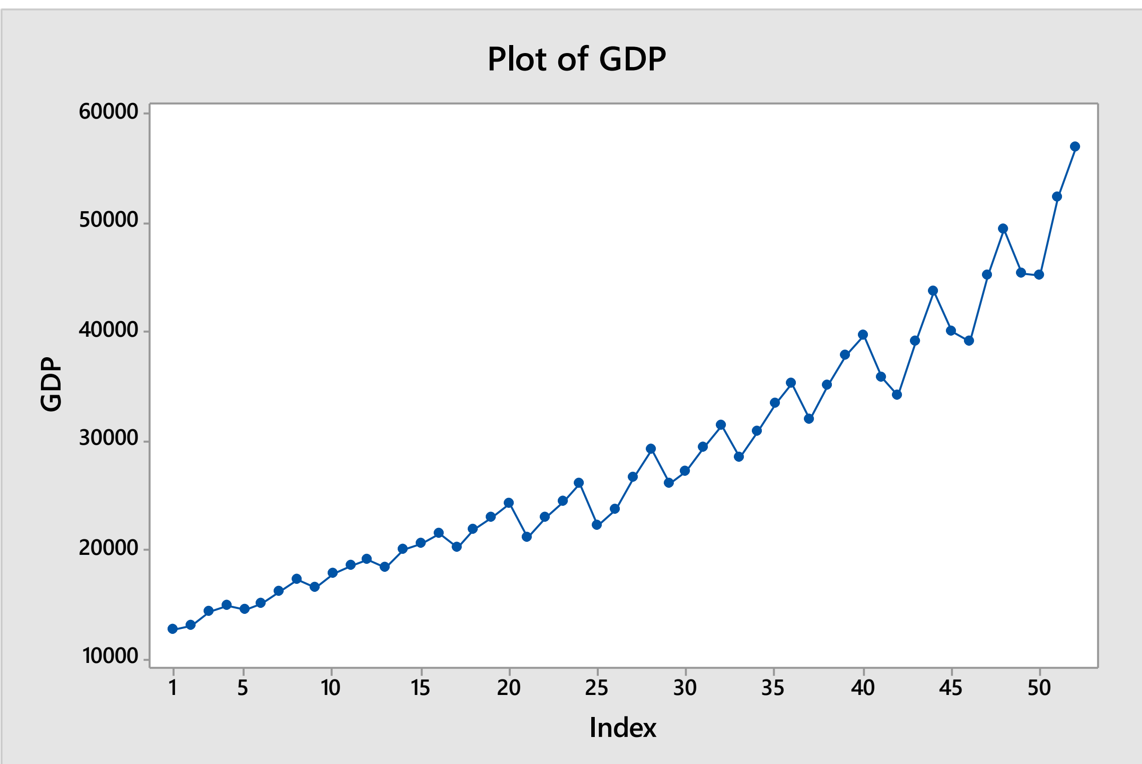


Figure 1: Graphs showing the GDP and Nigerian Crude Oil Production

ISSN NUMBER: 1116-249X

3.2 The MIDAS Approach

Mixed data sampling (MIDAS) regression is a flexible class of time series models that incorporates variables sampled at different frequencies into one regression model. It is usually characterized by response and predictor variables sampled at lower and higher frequencies respectively Clements and Galavao (2006). Specifically, the model under consideration is:

$$y_t = X_t^{'}\beta + f\left(\left\{X_{t/S}^{H}\right\}, \theta, \lambda\right) + \in_t \tag{1}$$

Where

- y_t is the dependent variable, sampled at a low frequency, at date t
- X_t is the set of regressors sampled at the same low frequency as y_t
- $X_{t/S}^{H}$ is a set of regressors sampled at a higher frequency with S values for each low frequency value. Note that $X_{t/S}^{H}$ is not restricted to the S values associated with the current t as it may include values corresponding to lagged low frequency values.
- f is a function describing the effect of the higher frequency data in the lower frequency regression.
- β, λ *and* θ vectors of parameters to be estimated.

In contrast, MIDAS estimation offers several different weighting functions which may occupy the middle ground between the unrestricted and the equally weighted aggregation approaches.

The MIDAS weighting functions reduce the number of parameters in the model by placing restrictions on the effects of high frequency variables at various lags.

3.2.1 Step Weighting

The simplest weighting method employs the step function:

$$y_t = X_t^{'}\beta + \sum_{\tau=0}^{k-1} X_{(t-\tau)/S^{'}\varphi\tau}^{H} + \in_t \tag{2}$$

where

- k is a chosen number of lagged high frequency periods to use (where k may be less than or greater than S
- η is a step length
- $\varphi_m = \theta_i \ \ for\ k = \text{int}(m/\eta)$

3.2.2 Almon (PDL) Weighting

Almon lag weighting (also known as polynomial distributed lag or PDL weighting) is widely used to place restrictions on lag coefficients in autoregressive models, which is a natural candidate for the mixed frequency weighting. For any high frequency lag up to k, the regression coefficients are modeled as a dimensional lag polynomial in the MIDAS parameters . We shall write the resulting restricted regression model as:

$$y_t = X_t^{'}\beta + \sum_{\tau=0}^{k-1} X_{(t-\tau)/S^{'}}^{H}\left(\sum_{j=0}^{p}\tau^{j}\theta_j\right) + \in_t \tag{3}$$

ISSN NUMBER: 1116-249X

where p is the Almon polynomial order, and the chosen number of lags k may be less than or greater than S. The number of coefficients that will be estimated depends on the polynomial order and not the number of high frequency lags.

3.2.3 Beta Weighting

The normalized Beta weighting approach is given as:

$$y_t = X_t^{/}\beta + \sum_{r=0}^{k-1} X_{t-r}^{H}\left(\frac{\omega_r^{\theta_1-1}(1-\omega_r)^{\theta_2-1}}{\sum_{j=0}^{k}\omega_j^{\theta_1-1}(1-\omega_j)^{\theta_2-1}} + \theta_3\lambda + \varepsilon_t\right) \tag{4}$$

Where k is a number of lags, λ is a slope coefficient that is common across lags, and

$$\omega_i = \begin{cases} \delta & i=0 \\ i/(k-1) & i=1,\ldots,k-2 \\ 1-\delta & i=k \end{cases} \tag{5}$$

where δ is a small number

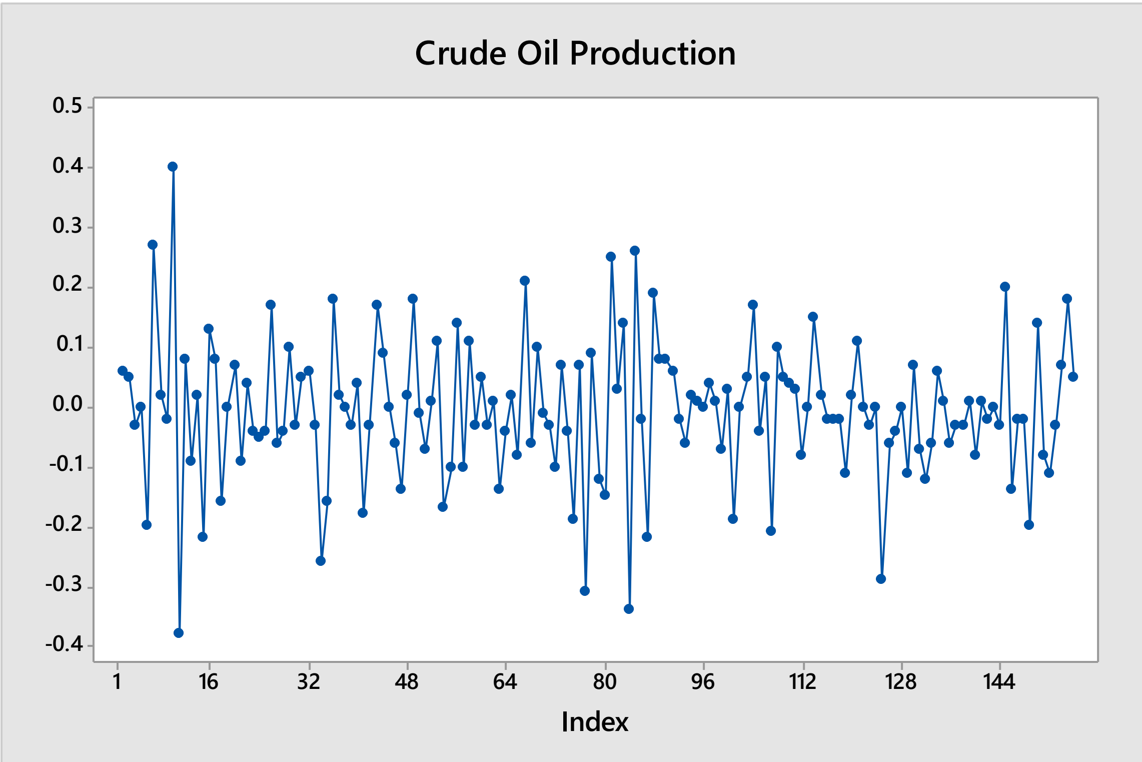
where k is a chosen number of lags, λ is a slope coefficient that is common across lags, and the differential response comes via the exponential weighting function and the lag polynomial which depends on the two MIDAS coefficients θ_1 and θ_2 .

3.3 Empirical Analysis, Results and Discussion

In the MIDAS regression model, the assumption of stationary data must be made. One way to cope with non-stationary data is to do a transformation, so it is expected to obtain stationary data so that the analysis of the next step can be done.

From Fig. 2, it can be concluded that the data structure of GDP and Crude Oil Production over time has constant fluctuations of data, and the fluctuations in the data are around the mean value. Therefore, the data is considered stationary.

ISSN NUMBER: 1116-249X



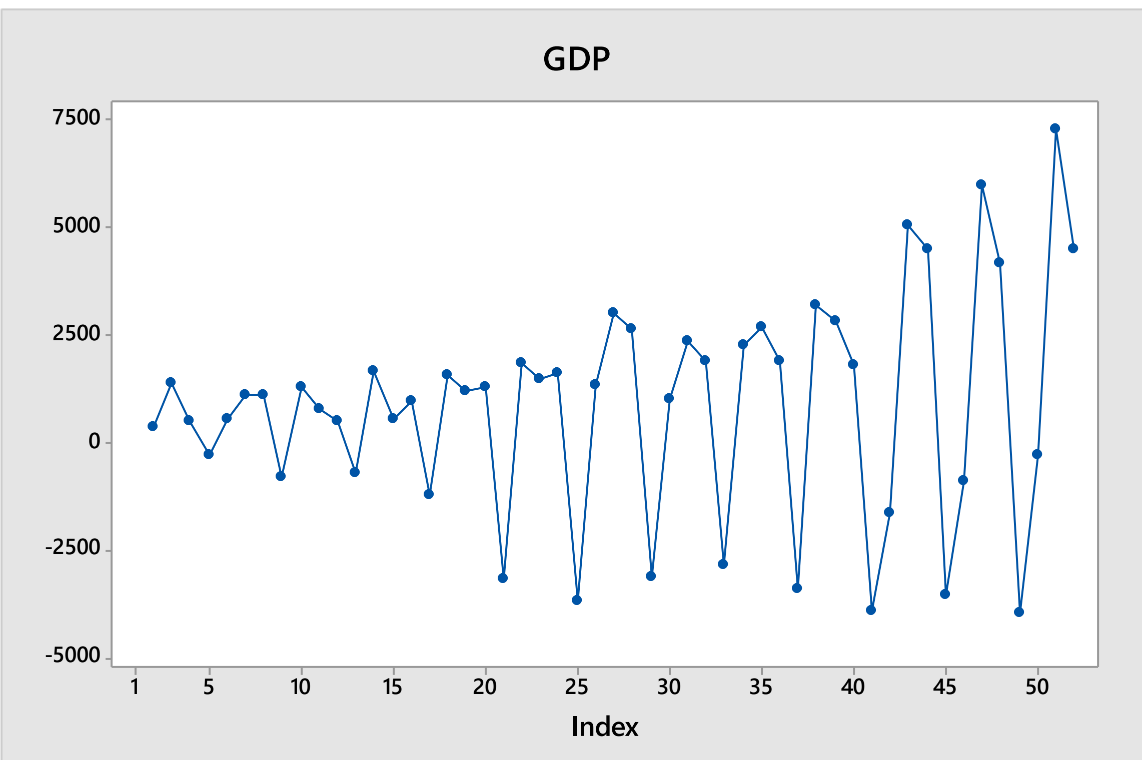


Figure 2: Nigerian GDP and Crude Oil Production

ISSN NUMBER: 1116-249X

Table 1: The results of the MIDAS estimation

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	184.3235	64.73924	2.847167	0.0066
DGDP(-1)	0.849895	0.054631	15.55701	0.0000
Page: Crude Oil Production Series: Crude Oil Production (-1) Lags: 4				
PDL01	17.67392	59.25475	0.298270	0.7668
PDL02	-36.04685	70.38194	-0.512161	0.6110
PDL03	10.61885	17.67916	0.600642	0.5510
R-squared	0.966840	Mean dependent var		1018.531
Adjusted R-squared	0.966164	S.D. dependent var		38.28671
S.E. of regression	7.042723	Akaike info criterion		6.897940
Sum squared resid	2430.397	Schwarz criterion		7.087335
Log likelihood	-170.8975	Hannan-Quinn criter.		6.970313
Durbin-Watson stat	2.139339			
COPM(-1)	Lag	Coefficient	Distribution	
	0	-7.754083	*	
	1	-11.94437	*	
	2	5.103041	*	

Here the top portion of the output describes the estimation sample, MIDAS method, and other estimation settings. The **Exponential Almon** weighting with restricted endpoints $\theta_3 = 0$ .The first section displays coefficients, standard errors, and *t*-statistics for the low frequency regressors. These results show standard regression output. The coefficient results for the common SLOPE coefficient (λ) and the free MIDAS beta weight coefficients (θ) are displayed directly below.

The lag value of the dependent variable GDP is statistically significant 0.849895(Prob. 0.0000).

The actual lag coefficients are obtained by applying weights to this overall slope. The shape of the weight function is determined by the remaining MIDAS coefficients. The (θ_1) coefficient, labeled EXPPDL01, is somehow close to 1, so that the lag pattern depends primarily on EXPPDL02 (θ_2).

The low negative estimate of (θ_2), is statistically different from 1, and value of 10.61885, implies that the lag pattern is slowly increasing as shown in the lag coefficient graph at the bottom of the output. We conclude that the coefficient that the zero die frequency lag of Crude Oil Production has a large impact on real GDP, but the effect did not dies off pretty quickly. The endpoint coefficient has been restricted to be zero so that it does not appear in the output.

In this empirical study, the parameter estimation for the MIDAS regression model with an exponential Almon function as well as with a PDL/Almon is done. The optimal lag

ISSN NUMBER: 1116-249X

determination is based on the smallest values of the AIC and SC. The values of the AIC and SC for the MIDAS regression model are as follows.

Table 2. Value of AIC and BIC Model of MIDAS Regression PDL

Lag	AIC	SC
Lag 4	6.897940	7.087335
Lag 9	6.925461	7.118504

Table 3. Value of AIC and SC Model of MIDAS Regression BETA

Lag	AIC	SC
Lag 4	6.926010	7.153283
Lag 9	6.960522	7.192174

Based on the smallest AIC and SC values in Table 2, it is found that the MIDAS regression PDL with the optimal lag is the model with lags 4. While the optimal lags in the MIDAS regression Beta model in Table 3, is at Lag 4.

Table 4. Value of RMSE

MODEL	**RMSE**
MIDAS BETA	9.036120
MIDAS PDL	8.939251
Step	28.97003

From the table 4 above, it is observed that based on the Root Mean Square Error the MIDAS Almon (PDL) regression model provided a better model estimation than the MIDAS Step weighting and the MIDAS Beta.

Hence, the MIDAS Almon PDL regression model can be used to forecast the Nigerian GDP growth rate.

4. Forecasting Nigerian GDP

In the forecasting stage, the Nigerian GDP for the next thirty-two steps from 2015Q1 to 2022Q4 will be forecasted.

ISSN NUMBER: 1116-249X

.

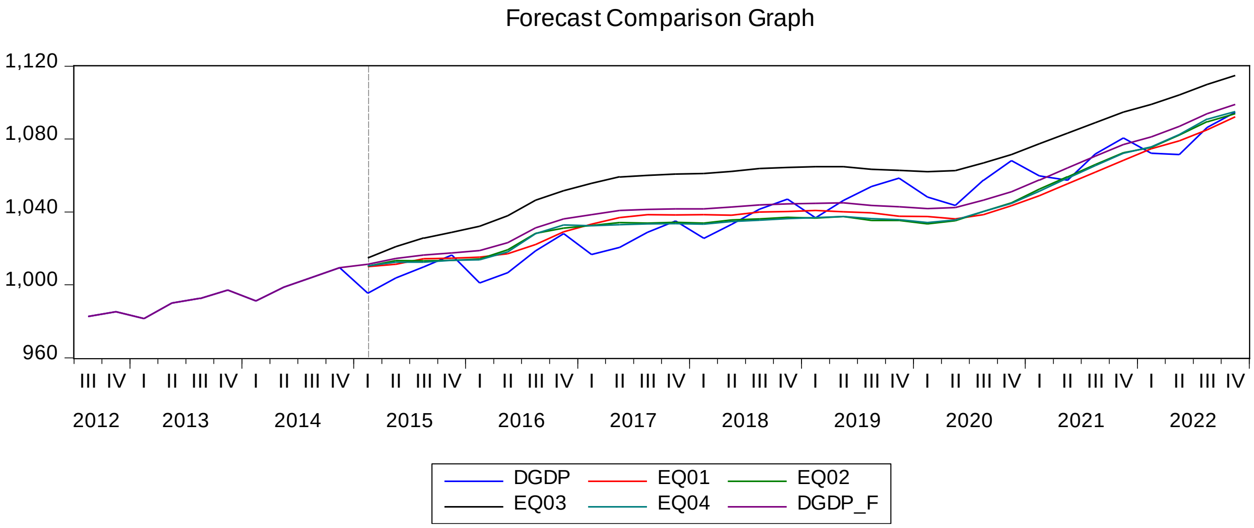


Figure 5: Forecast Comparison Graph

The upward trend in the forecast graph suggests a positive economic outlook and highlights the potential usefulness of the MIDAS model for economic forecasting and policy-making.

5. CONCLUSION

In this study, an attempt was made to investigate the Nigeria' Crude Oil Production and the Nigeria Gross Domestic Product. It was observed that based on the Root Mean Square Error the MIDAS Almon (PDL) regression model provided a better model estimation than the MIDAS Beta and the MIDAS Step weighting. The monthly Crude oil production has a positive effect on GDP as the slope coefficient is statistically significant 0.849895(Prob. 0.0000). It is therefore important for appropriate policy formulation and implementation of such policies to encourage and boost these variables for effective management of Crude oil production in Nigeria. Hence, direct relationship between Crude oil production and GDP is needed to diversify the economy base to enhance productive activities in Nigeria and better crude oil production. Audrone *et al.*, (2023), while their study focuses on financial variables, the idea of using MIDAS to capture relationships between high-frequency and low-frequency variables aligns with this study.

REFERENCES

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models", CIRANO Working Papers 2004s-20, CIRANO.

Clements, M. P. and Galavao, A. B. (2006). Macroeconomic Forecasting with mixed Frequency Data: Forecasting US output growth and inflation. *Warwick Economic Research Paper No. 773*. University of Warwick.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006), "Predicting volatility: getting the most out of return data sampled at different frequencies", *Journal of Econometrics*, 131, 59–95.

ISSN NUMBER: 1116-249X

Tsui, A., Xu, C., and Zhang, Z. (2013). Forecasting Singapore economic growth with mixed-frequency data. Proceedings of International Congress on Modelling and Simulation (pp. 1399-1405).

Adeniji S. O., Timilehin J. O. and Gamaliel O. E. (2017). The Impact of Money Supply on Nigeria Economy: A Comparison of Mixed Data Sampling (MIDAS) and ARDL Approach. *Euroeconomica* Issue 2(36).

Bonnier, J.-B. (2022), 'Forecasting crude oil volatility with exogenous predictors: As good as it GETS?', Energy Economics 111, 106059.

Acha, K., Amalahu, C. C. and Eziokwu, C. (2023).Modeling and Forecasting of Nigeria Crude Oil Production. ***Journal of Mathematics and Statistics Studies***. 4(1): 58-67

Audrone V., Hoang N. and Minh-N. (2023). Bayesian Predictive Distributions of Oil Returns Using Mixed Data Sampling Volatility Models. ISSN 1403-0586

Hassan, U., & Ismail, M. T. (2023). Improving forecasting accuracy using quantile regression neural network combined with unrestricted mixed data sampling. *Journal of the Nigerian Society of Physical Sciences*, 5(4), 1394

Lyu, Y., Qin, F., Ke, R., Yang, M., & Chang, J. (2024). Forecasting the VaR of the crude oil market: A combination of mixed data sampling and extreme value theory. *Energy Economics*, 133, 107500.

ISSN NUMBER: 1116-249X