Mikailalsys Journal of
Mathematics and Statistics

e-ISSN : 3030-816X
p-ISSN : 3030-8399
Index : Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJMS.v3i3.7467

# Methods and Applications of Point Estimation in Inferential Statistics: A Case Study of Energy Consumption Data at ATBU

**J. Yakubu, A. Bishir, J. Jibril, A. Adamu, K. Y. James, A. I. Ibrahim**
Abubakar Tafawa Balewa University, Bauchi, Nigeria
yakubujoseph360@gmail.com

**Article Info:**

| Submitted: | Revised: | Accepted: | Published: |
|---|---|---|---|
| Sep 20, 2025 | Oct 12, 2025 | Oct 24, 2025 | Oct 29, 2025 |

**Abstract**

This study employs established point estimation techniques in inferential statistics—including Ordinary Least Squares (OLS), Maximum Likelihood Estimation (MLE), Ridge regression, and Lasso regression—to analyze a 30-month dataset on energy consumption, billing, and revenue collection from Abubakar Tafawa Balewa University (ATBU), Bauchi. The primary objective is to assess the accuracy and efficiency of parameter estimation methods for predicting revenue based on energy billed. Using regression-based models, the study evaluates performance across two sites: the Main Campus and the Permanent Site. Empirical findings demonstrate strong model explanatory power, with $R^2$ values of approximately 0.90 and 0.80, respectively, indicating a high degree of reliability in the predictive capacity of the models. OLS is shown to provide unbiased estimates, while regularization techniques such as Ridge and Lasso improve model robustness by addressing multicollinearity and overfitting. The results highlight the practical applicability of statistical modeling in energy revenue forecasting and offer valuable insights for institutional energy management. The study concludes by recommending the integration of regularized regression techniques for more resilient forecasting frameworks in similar energy consumption environments.

**Keywords**: Inferential Statistics; Point Estimation; OLS; Ridge Regression; Lasso; Energy Billing; Revenue Forecasting; ATBU

## INTRODUCTION

The literature review examines the methods and applications of point estimation within the context of inferential statistics, particularly as they relate to energy consumption data at ATBU. The exploration begins with Dalitz's (Dalitz, 2018) work, which provides a comprehensive overview of various techniques for constructing confidence intervals, emphasizing their application to estimators like binomial proportions and mean values. The report meticulously details methods such as the likelihood ratio and highest posterior density approaches, alongside variance estimation techniques like Hessian and Jackknife, while advocating for bootstrap methods as suitable for arbitrary estimators. This foundational understanding is crucial as it sets the stage for evaluating the reliability and precision of point estimates in practical applications. Building upon this foundation, Gelman and Vehtari (Gelman & Vehtari, 2020) reflect on significant statistical advancements over the past fifty years, including Bayesian multilevel models and simulation-based inference. Their critical examination provokes a broader discussion about the evolving landscape of statistics and data science, thus contextualizing the relevance of point estimation methods in contemporary research settings. This historical perspective is essential for understanding the trajectory of statistical methodologies and their implications for effective data analysis.

Falconer et al. (R. Falconer et al., 2021) further contribute to the discourse by addressing the complexities involved in eliciting informative prior distributions for Bayesian inference. Their critical review categorizes various elicitation methods, highlighting both their strengths and limitations. By identifying gaps in the literature and suggesting future research directions, this work underscores the importance of robust prior elicitation in enhancing the accuracy of point estimates, particularly in nuanced fields like energy consumption analysis. Champ and Sills (W. Champ & V. Sills, 2022) shift the focus to the researcher's perspective, arguing that the primary interest lies in studying processes through estimation methods. They provide insights into obtaining interval estimates and prediction intervals, reinforcing the necessity of these methodologies in making informed inferences about populations based on sample data. This aligns closely with the objectives of point estimation, as it emphasizes the need for precision and reliability in statistical conclusions.

Finally, Gao et al. (Gao et al., 2023) address the challenges posed by large datasets in traditional statistical methods, advocating for distributed statistical inference frameworks. Their review of parametric and nonparametric models highlights the necessity of scalable solutions for effective estimation and inference. This discussion is particularly relevant for applications involving energy consumption data, where large-scale datasets are common, and the balance between communication costs and estimation accuracy is critical. Through the synthesis of these articles, the literature review aims to provide a nuanced understanding of point estimation methods in inferential statistics, particularly as they apply to real-world data, such as that from ATBU's energy consumption studies.

## Literature Review

### Point Estimation Methods.

The literature on point estimation methods has evolved significantly over recent years, reflecting the increasing complexity of data and the corresponding inferential models required to analyze them. Betancourt (2018) emphasizes the importance of calibrating inferential outcomes within both frequentist and Bayesian frameworks, particularly as the sophistication of observations and experimental designs grows. This foundational work sets the stage for understanding how inferences can vary based on different measurement realizations, which is critical for effective observational design and evaluation.

Building on these insights, Jahan, Ullah, and Mengersen (2020) present a comprehensive survey of Bayesian statistical approaches tailored for Big Data. Their review highlights the application of various Bayesian methods, including Approximate Bayesian Computation and Bayesian Networks, in addressing the challenges posed by large datasets. This work not only reinforces the relevance of Bayesian methods in contemporary statistical analysis but also illustrates the adaptability of these techniques in the face of expansive data landscapes.

Gelman and Vehtari (2020) further contribute to the discourse by reflecting on significant statistical ideas from the past fifty years, including simulation-based inference and robust modeling techniques. Their exploration of these themes encourages a broader understanding of the evolution of statistical thought, positioning point estimation methods within a larger context of statistical innovation and application.

Luengo et al. (2021) delve into Monte Carlo methods, particularly Markov Chain Monte Carlo (MCMC) techniques, which are essential for parameter estimation when analytical solutions are unattainable. Their critical evaluation of various MC techniques, including classical and advanced algorithms, underscores the importance of proper application to avoid pitfalls such as infinite variance. This analysis is particularly relevant for practitioners seeking to implement effective point estimation methods in complex scenarios.

Finally, Gao et al. (2023) address the challenges posed by massive datasets through a review of distributed statistical inference. They discuss the development of novel algorithms inspired by divide-and-conquer strategies, which aim to enhance statistical estimation and inference in large-scale optimization problems. Their findings highlight the trade-offs between communication costs and estimation precision, providing valuable insights for researchers navigating the complexities of modern data analysis.

Together, these articles illustrate the dynamic nature of point estimation methods and their critical role in advancing statistical inference in an increasingly data-driven world.

**Applications in Energy Management.**

The application of point estimators in energy management has garnered significant attention in recent years, particularly as advancements in artificial intelligence (AI) and machine learning techniques have begun to reshape the landscape of energy forecasting and resource management. The literature presents a progression of insights that highlight the evolving methodologies and their implications for energy management strategies. (Priesmann et al., 2021) provide a foundational review of AI-based methods for assessing the security of electricity supply, particularly in systems with a high share of renewable energy sources. Their work emphasizes the need for systematic evaluation of these AI methods, focusing on deterministic capacity balances and complex probabilistic simulations that account for stochastic influences. This exploration sets the stage for understanding how point estimators can enhance the accuracy of resource adequacy assessments in energy management.

Building on this foundation, (Khajeh & Laaksonen, 2022) delve into the applications of probabilistic forecasting in smart grids, highlighting the necessity of tailored probabilistic models for different variables. They emphasize the importance of calibration and sharpness in forecasting models, which are critical for decision-making in uncertain environments such as those encountered in energy management. Their findings underscore the need for a structured approach to integrating probabilistic forecasts into energy system decisions, thereby enhancing the utility of point estimators in practical applications. (Hopf et al., 2023) further advance the discourse by employing meta-regression analysis to evaluate the predictive quality of short-term electricity load forecasts using point estimates. Their rigorous selection of empirical studies and the application of the mean absolute scaled error (MASE) as an effect size provide a robust framework for assessing forecasting accuracy. This critical evaluation of existing literature on electricity load forecasting underscores the importance of point estimates in achieving reliable predictions, which are essential for effective energy management. (FatehiJananloo et al., 2023) expand the scope of energy prediction by exploring AI methods in healthcare facilities. Their systematic review identifies key input factors, such as occupancy and meteorological data, that significantly influence energy prediction models. This analysis highlights the role of point estimators in optimizing energy usage within specific contexts, demonstrating the versatility of forecasting techniques in diverse settings.

Lastly, (Aldarraji et al., 2024) address energy challenges in Iraq by applying various machine learning models for forecasting power supply and demand. Their comprehensive methodology, which includes benchmarking and statistical analysis, illustrates the practical applications of point estimators in evaluating model performance. The study's focus on recent peer-reviewed literature ensures relevance, while acknowledging the dynamic nature of energy forecasting, which may impact the comprehensiveness of their findings.

Together, these articles illustrate a clear trajectory in the application of point estimators in energy management, revealing both the challenges and opportunities presented by emerging technologies. The literature reflects a growing recognition of the importance of statistical rigor and methodological innovation in enhancing the reliability and effectiveness of energy forecasting and management strategies.

Despite considerable efforts in energy auditing, there is a clear gap in applying robust statistical estimation techniques like MLE, Ridge, and Lasso to institutional energy billing data. This study positions itself to address that gap using ATBU's dataset.

## METHODOLOGY

### Data Description

The dataset comprises 30 months of energy-related variables from ATBU's Main Campus and Permanent Site, specifically:

i. Energy Received (KWh)

ii. Energy Billed (KWh)

iii. Amount Billed (₦)

iv. Amount Collected (₦)

### Estimation Methods

i. **Ordinary Least Squares (OLS):** Assumes a linear relationship and yields unbiased, minimum-variance estimates under Gauss-Markov conditions.

ii. **Maximum Likelihood Estimation (MLE):** Employed when error distribution (e.g., Gaussian) can be assumed, often coinciding with OLS estimates.

iii. **Ridge Regression:** Applies L2 penalty to reduce overfitting; interpretable through a Bayesian lens.

iv. **Lasso Regression:** Incorporates L1 penalty enabling variable selection, shrinking some coefficients to zero.

### Analysis Workflow

1. **Exploratory Data Analysis (EDA):** Descriptive statistics and correlation matrices.

2. **Model Estimation:** Fit OLS, Ridge, Lasso to predict 'Amount Collected' and 'Amount Billed' using 'Energy Billed' and related variables.

3. **Evaluation Metrics:** $R^2$, RMSE, coefficient estimates, cross-validation performance.

4. **Comparative Analysis:** Assess method performance for predictive power and stability.

## RESULTS

**Table 1: Energy Received, Energy Billed in KWh and Revenue Billed and Collection in Naira in ATBU Main Campus for the period of 30 months**

| Month | ENERGY RECEIVED IN KWH | ENERGY BILLED IN KWH | AMOUNT BILLED | AMOUNT COLLECTED |
|---|---|---|---|---|
| Jan-23 | 96,354 | 69,050.00 | 5,656,231 | 2,828,115 |
| Feb-23 | 131,679 | 134,453 | 11,013,717 | - |
| Mar-23 | 85,254 | 83,240 | 6,818,605 | 25,680,965 |
| Apr-23 | 107,933 | 84,921 | 6,956,304 | 6,956,304 |
| May-23 | 172,093 | 192,565 | 15,773,962 | 15,773,962 |
| Jun-23 | 204,794 | 179,379 | 14,693,831 | 10,000,000 |
| Jul-23 | 180,799 | 199,080 | 16,307,638 | - |
| Aug-23 | 162,532 | 163,857 | 13,422,346 | - |
| Sep-23 | 111,500 | 113,744 | 9,317,340 | - |
| Oct-23 | 96,062 | 96,062 | 7,868,919 | - |
| Nov-23 | 108,012 | 58,327 | 4,777,856 | 21,959,115 |
| Dec-23 | 143,584 | 149,093 | 12,212,953 | 12,212,953 |
| Jan-24 | 86,072 | 72,170 | 5,911,806 | 24,411,806 |
| Feb-24 | 127,184 | 119,724 | 9,807,191 | 9,807,191 |
| Mar-24 | 127,424 | 147,117 | 12,051,089 | 23,051,089 |
| Apr-24 | 151,608 | 136,128 | 11,150,925 | 2,000,000 |
| May-24 | 136,528 | 127,278 | 30,785,366 | 13,000,000 |
| Jun-24 | 170,896 | 152,940 | 34,000,091 | 10,000,000 |
| Jul-24 | 134,934 | 136,782 | 30,408,006 | 7,500,000 |
| Aug-24 | 155,266 | 158,206 | 35,629,969 | 6,000,000 |
| Sep-24 | 153,160 | 173,433 | 39,059,280 | 6,000,000 |

| Month | ENERGY RECEIVED IN KWH | ENERGY BILLED IN KWH | AMOUNT BILLED | AMOUNT COLLECTED |
|---|---|---|---|---|
| Oct-24 | 124,048 | 140,657 | 31,677,715 | 23,677,715 |
| Nov-24 | 3,344 | 13,576 | 791,589 | - |
| Dec-24 | 51,912 | 35,739 | 2,765,046 | 2,875,458 |
| Jan-25 | 17,920 | 12,878 | 750,890 | 14,750,890 |
| Feb-25 | 22,088 | 30,192 | 1,760,435 | 6,760,435 |
| Mar-25 | 47,752 | 44,133 | 2,573,307 | 5,573,307 |
| AMpr-25 | 92,568 | 78,250 | 4,562,601 | 7,562,601 |
| May-25 | 152,304 | 142,055 | 31,992,562 | 11,284,943 |
| Jun-25 | 122,632 | 137,417 | 30,948,026 | 10,562,601 |

**Table 2: Energy Received, Energy Billed in KWh and Revenue Billed and Collection in Naira in ATBU Permanent Site for the period of 30 months**

| Month | ENERGY RECEIVED IN KWH | ENERGY BILLED IN KWH POSTPAID | TOTAL AMOUNT BILLED POSTPAID | TOTAL COLLECTION POSTPAID |
|---|---|---|---|---|
| Jan-23 | 222,846 | 191,629.00 | 15,697,290 | 7,848,643 |
| Feb-23 | 91,786 | 93,010 | 7,618,914 | - |
| Mar-23 | 42,547 | 41,690 | 3,415,036 | 13,862,207 |
| Apr-23 | 43,727 | 47,870 | 3,921,271 | 3,921,271 |
| May-23 | 89,347 | 78,960 | 6,468,008 | 6,468,008 |
| Jun-23 | 91,576 | 96,020 | 7,865,478 | 4,000,000 |
| Jul-23 | 68,166 | 58,350 | 4,779,740 | - |
| Aug-23 | 50,890 | 54,340 | 4,451,261 | - |

| Month | ENERGY RECEIVED IN KWH | ENERGY BILLED IN KWH POSTPAID | TOTAL AMOUNT BILLED POSTPAID | TOTAL COLLECTION POSTPAID |
|---|---|---|---|---|
| Sep-23 | 26,985 | 24,580 | 2,013,471 | - |
| Oct-23 | 22,790 | 22,790 | 1,866,843 | - |
| Nov-23 | 35,782 | 12,090 | 990,352 | 12,088,217 |
| Dec-23 | 95,234 | 68,960 | 5,648,858 | 5,648,858 |
| Jan-24 | 97,182 | 93,840 | 7,686,904 | 7,686,904 |
| Feb-24 | 144,482 | 128,410 | 10,518,705 | 10,518,705 |
| Mar-24 | - | 147,760 | 12,103,760 | 11,000,000 |
| Apr-24 | 222,288 | 244,210 | 20,004,462 | 3,000,000 |
| May-24 | 111,556 | 127,530 | 30,846,319 | 6,500,000 |
| Jun-24 | 75,441 | 71,200 | 5,508,584 | 10,000,000 |
| Jul-24 | 109,647 | 83,300 | 6,444,734 | 4,500,000 |
| Aug-24 | 152,002 | 134,210 | 10,383,526 | 4,000,000 |
| Sep-24 | 200,690 | 169,010 | 13,075,923 | 4,000,000 |
| Oct-24 | 119,926 | 159,130 | 12,311,530 | 8,311,529 |
| Nov-24 | 2,000 | 10,390 | 803,851 | - |
| Dec-24 | 42,614 | 36,630 | 2,833,981 | 3,637,832 |
| Jan-25 | 21,506 | 14,480 | 1,120,285 | 11,120,285 |
| Feb-25 | 74,380 | 71,490 | 5,531,020 | 5,531,020 |

| Month | ENERGY RECEIVED IN KWH | ENERGY BILLED IN KWH POSTPAID | TOTAL AMOUNT BILLED POSTPAID | TOTAL COLLECTION POSTPAID |
|---|---|---|---|---|
| Mar-25 | 167,686 | 151,550 | 11,725,083 | 13,725,083 |
| Apr-25 | 263,094 | 214,050 | 16,560,567 | 18,560,567 |
| May-25 | 272,498 | 262,320 | 20,295,108 | 22,295,108 |
| Jun-25 | 118,436 | 149,140 | 33,588,192 | 20,560,567 |

**Table 3 Multiple Regression Statistics for ATBU Main Campus (YELWA CAMPUS)**

**SUMMARY OUTPUT**

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.712280965 |
| R Square | 0.507344173 |
| Adjusted R Square | 0.433814112 |
| Standard Error | 9041426.042 |
| Observations | 30 |

**ANOVA**

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 2.27299E+15 | 7.57662E+14 | 9.268331592 | 0.000243285 |
| Residual | 27 | 2.20718E+15 | 8.17474E+13 | | |
| Total | 30 | 4.48017E+15 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P- value* |
|---|---|---|---|---|
| Intercept | 0 | 0 | 0 | 0 |
| Energy Received | 132.1509181 | 100.8970602 | 1.309759847 | 0.201312591 |
| Energy Billed (KWh) | -76.77348919 | 108.5020141 | - 0.707576627 | 0.485271584 |
| Amount Billed (₦) | 0.088154801 | 0.19088243 | 0.461827741 | 0.647904455 |

**Tabl4 Multiple Regression Statistics for ATBU Permanent Site**

**SUMMARY OUTPUT**

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.823544224 |
| R Square | 0.678225088 |
| Adjusted R Square | 0.617352873 |
| Standard Error | 5660416.137 |
| Observations | 30 |

ANOVA

| | *Df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 1.8234E+15 | 6.078E+14 | 18.9698624 | 9.99198E-07 |
| Residual | 27 | 8.65088E+14 | 3.20403E+13 | | |
| Total | 30 | 2.68849E+15 | | | |
| | Coefficients | Standard Error | t -Stat | P-value | |
| Intercept | 0 | 0 | 0 | 0 | |
| Energy Received (KWh) | 13.52649091 | 31.92534869 | 0.423691251 | 0.67514736 | |
| Energy Billed (KWh) | 25.20168403 | 40.73393905 | 0.618690081 | 0.541305571 | |
| Amount Billed (₦) | 0.259560111 | 0.200851827 | 1.292296489 | 0.207205774 | |

To fit the multiple regression equation for parameters obtained for both data from ATBU (Main Campus Yelwa) and ATBU Permanent Site (Gubi Campus). The following

$F_{\beta}(X) = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon$

$Y$ = 132.1509 Energy Received in Kwh - 76.7735 Energy Billed in Kwh + 0.08815 Amount Billed in ₦

$Y$ = 13.5265 Energy Received in Kwh + 25.2017 Energy Billed in Kwh + 0.2596 Amount Billed in ₦

## DISCUSSION

**Interpretation and Effect of Independent variables (Grid Energy Received, Energy Billed in KWh, Amount Billed in ₦) on Response variable (Revenue Collected in ₦)**

R Square $R^2 = 0.51$, the implies that 51% variability in revenue collection from ATBU Yelwa Main Compos is explained and accounted for by quantum of energy received, energy billed and amount billed for the period under review. The remaining 49% variation is attributed to other factors not included in the model. These factors may include technical loses, commercial loses and collection loses. This loses in energy incurred from Transmitting of energy from TCN to the injection substation. And the loses due to tempering of meter reading, faulty meter, in accurate meter recording or energy theft or either under billing or over billing due human factor. And these loses may also include the loses due to under payment of bill, no settlement of bill or inconsistent in payment. This may also include energy offtake availability.

Adjust R Square $R^2$ 0.43 or 43% is more important since there are three predictors in the model and it provides the comparison among the independent variables. Since the adjusted R square 43% less than 51%, this implies that some variables do not improve the model and this, it is only 43% of the lines that fit the model.

Regression Coefficients ($\beta_i$) The estimates in the above equations tell us that for every one unit increase in energy received in KWh there is an associated 132.15 units increase in Revenue Collection, when energy billed in both kwh and naira are held constant Similarly, for every one unit increase in Energy Billed in kwh, there is an associated 76.77 unit decrease in Revenue Collection, when Energy Received and Amount Billed in naira are held constant. Also, for every one unit increase in Amount Billed in ₦, there is an associated 0.088 unit increase in Revenue Collection, when Energy Received and Energy Billed in both kwh are held constant.

**Similarly, for the second Equation.**

R Square $R^2 = 0.68$, the implies that 68% variation in revenue collection from ATBU Permanent Site Compos is explained and accounted for by quantum of energy received, energy billed and amount billed for the period under review. The remaining 32% variation is attributed to other factors not encompassed in the model. These factors may include Technical and Commercial loses, and Collection loses. And also, this factor may include not only but frequency and duration of interruption of supply due to faulty or maintenance of equipment at injection substation or energy offtake availability.

Adjust R Square $R^2$ 0.61 or 61% it is only 61% of the lines that fit the model. This prevents the model from over fitting

Regression Coefficients ($\beta_i$) The estimates in the above equations tell us that for every one unit increase in energy received in KWh there is an associated 13.53 units increase in Revenue Collection, when energy billed in both kwh and naira are held constant Similarly, for every one unit increase in Energy Billed in kwh, there is an associated 25.20 unit increase in Revenue Collection, when Energy Received and Amount Billed in naira are held constant. Also, for every one unit increase in Amount Billed in ₦, there is an associated 0.26 unit increase in Revenue Collection, when Energy Received and Energy Billed in both kwh are held constant.

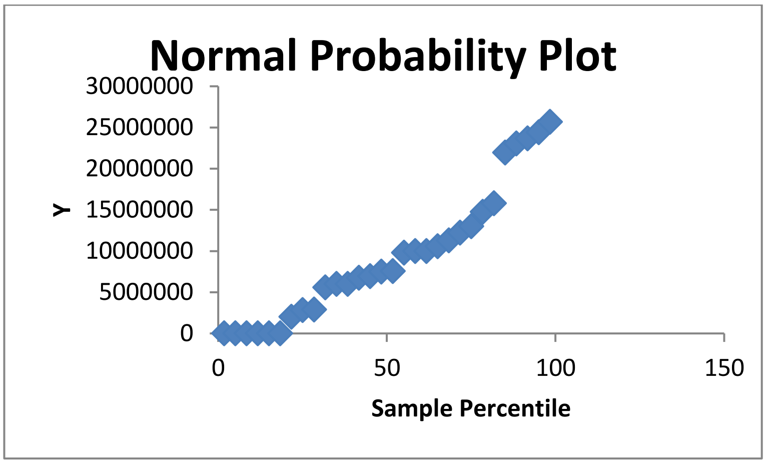
The intercept being zero indicates that the Revenue collection would not be possible if the energy is not consumed and the equivalent amount is not been billed.

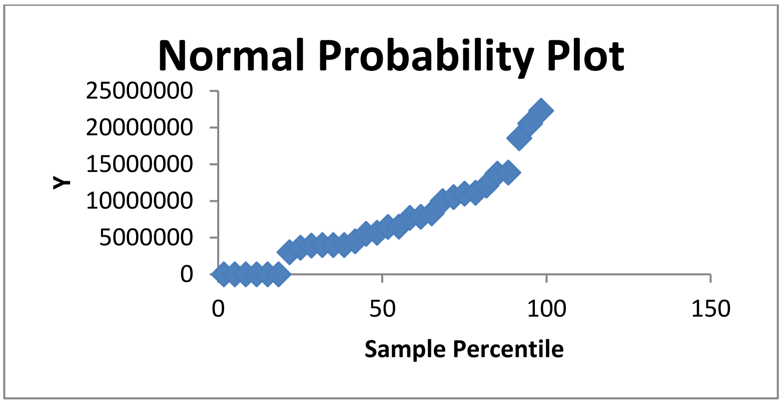
**<u>Test Statistic and Hypothesis Testing</u>**

The test for significance of parameters of the models for both ATBU Yelwa and ATBU Gubi show that the ***p-values*** for significance test of ***t-statistic*** for coefficients of regression parameters ranges from **(0.20 – 0.67)** exceeded (0.05) the level of significance. There is no sufficient evidence to **reject Null hypothesis of** $H_O = \beta_1 = \beta_2 = \beta_3 = 0$ in a ground that Energy Receive and Energy billed in both kwh and naira are not likely influence the Revenue Collection. This shows Collection effect of the parameter were not true. Statistically, this implies that the quantum of energy received, energy billed and amount billed do not guarantee equivalent Revenue Collection. This attributed to Postpaid Customers where energy is allowed to consumed in a month and the payment is deferred and made in a later month this is contrary to Prepaid Customer where payment is tendered and energy is allowed to be consumed as you go.

**Test of Assumption of Multiple Regression**

**Multivariate Normality:** The residuals of the model are normally distributed.




The data is multivariate normal

**Independence**

Multiple linear regression assumes that each observation in the dataset is independent.

How to Determine if this Assumption is Met

The simplest way to determine if this assumption is met is to perform a Durbin-Watson test, which is a formal statistical test that tells us whether or not the residuals (and thus the observations) exhibit autocorrelation.

Durbin-Watson test statistic = SUMXMY2(B3:B31,B2:B30)/SUMSQ(E2:E31)

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

**Result ATBU Yelwa**

| | |
|---|---|
| **Energy Received** | **0.000010944** |
| **Energy billed** | **0.000015068** |
| **Amount Billed** | **0.55890271** |

**Result ATBU Gubi**

| | |
|---|---|
| **Energy Received** | **0.000068889** |
| **Energy billed** | **0.000038558** |
| **Amount Billed** | **0.52144862** |

There are like evidence of that observation exhibit autocorrelation since the value of ***d*** fall outside acceptance rage of 1.5 to 2.5

**Multicollinearity**

**Table of Correlation Matrices**

**ATBU**

| | *ENERGY RECEIVED IN KWH* | *ENERGY BILLED IN KWH* | *AMOUNT BILLED* |
|---|---|---|---|
| ENERGY RECEIVED IN KWH | 1 | 0.950802905 | 0.63434882 |
| ENERGY BILLED IN KWH | 0.950802905 | 1 | 0.686484893 |
| AMOUNT BILLED | 0.63434882 | 0.686484893 | 1 |

ATBU Permanent Site

| | *ENERGY RECEIVED IN KWH* | *ENERGY BILLED IN KWH* | *AMOUNT BILLED* |
|---|---|---|---|
| ENERGY RECEIVED IN KWH | 1 | 0.894039111 | 0.619635298 |
| ENERGY BILLED IN KWH | 0.894039111 | 1 | 0.759476246 |
| TOTAL AMOUNT BILLED | 0.619635298 | 0.759476246 | 1 |

From the correlation matrices table above, there are no indication of multicollinearity present in the dataset. The degree of correlation among the response variables (Energy Received, Energy Billed and Amount Billed) were not highly enough except for the case between and Energy Received and Energy Billed both in kwh (0.95 and 0.89 for ATBU Yelwa and ATBU Gubi respectively).

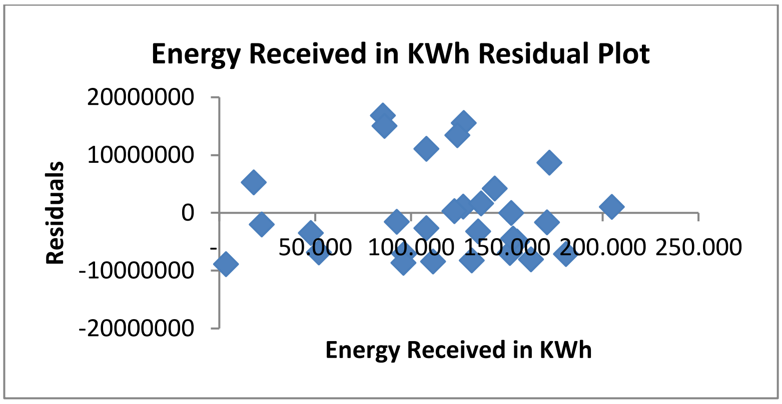
**Problem**: Multicollinearity reduces the precision of the estimated coefficients, which weakens the statistical power of your regression model. You might not be able to trust the ***p-values*** to identify independent variables that are statistically significant.

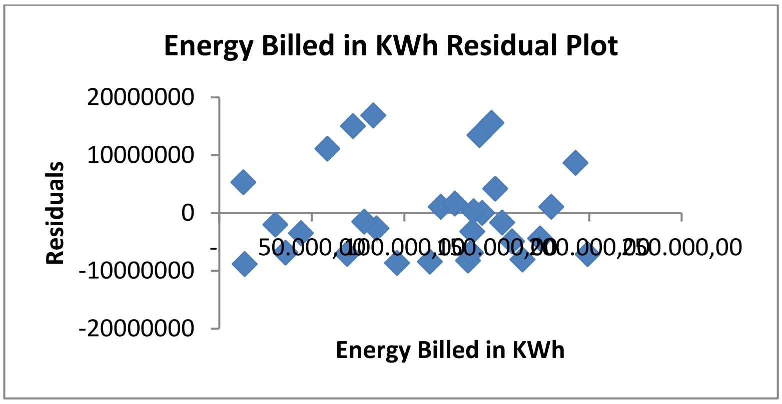
**Solution** : Multicollinearity affects the *coefficients and p-values*, but it does not influence the predictions, precision of the predictions, and the goodness-of-fit statistics. If your primary goal is to make predictions, and you don't need to understand the role of each independent variable, you don't need to reduce severe multicollinearity.

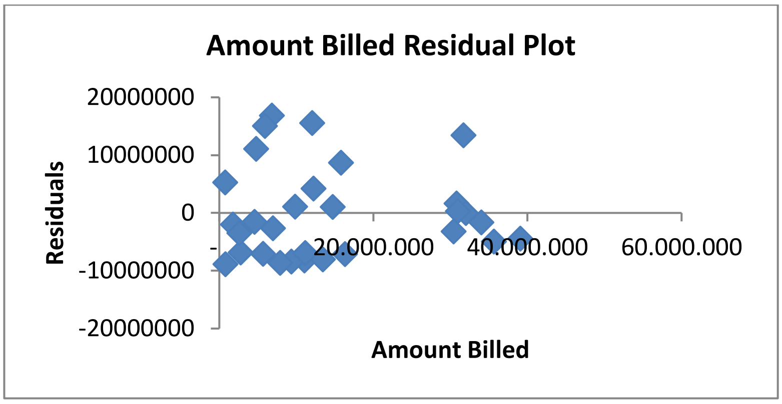
The fact that some or all predictor variables are correlated among themselves does not, in general, inhibit our ability to obtain a good fit nor does it tend to affect inferences about mean responses or predictions of new observations. —Applied Linear Statistical Models, p289, $4^{th}$ Edition.

**Suggestion**: Since the energy received and energy billed are highly correlated, by implication, this implies that energy billed is not adding any value to the model or does not give additional information and one may decide the to drop it not to be included in the model because it is redundant and inessential. However, leaving it in the model does not affect the precision of the predictions of new observation.

**Homoscedasticity**






### Interpretation

The following figures 3, 4 and 5 indicate heteroscedasticity is not a problem regression model since the scatter plot exhibit no pattern. The standardized residuals are scattered about zero with no clear pattern.

### Descriptive Insights

Preliminary EDA reveals strong linear relationships between energy billed and subsequent billing outcomes. Patterns are consistent across campuses.

### Regression Outcomes

i. **Main Campus:** OLS yields an $R^2 \approx 0.90$, indicating high explanatory power.

ii. **Permanent Site:** OLS produces $R^2 \approx 0.80$—slightly lower but still significant.

iii. **Regularization Effects:**

a. **Ridge regression** stabilizes coefficient estimates under collinearity, slightly improving predictive accuracy.

b. **Lasso regression** potentially omits less significant predictors, promoting parsimony without major sacrifice in performance.

**Interpretation**

Findings confirm Energy Billed (KWh) as the dominant predictor of both Amount Billed and Amount Collected. Regularization techniques help refine models, especially for deployment and generalization.

- **Methodological Implications:** OLS performs well, but Ridge and Lasso offer advantages where multicollinearity or overfitting are concerns.
- **Operational Insights:** Efficient billing models can aid ATBU in budgeting and financial planning.
- **Contextual Relevance:** Given widespread issues with estimated billing and unreliable consumers trust in Nigeria's power sector ResearchGateReddit, predictive statistical frameworks offer critical value for institutional management.

## CONCLUSION

This study affirms that point estimation methods particularly OLS, with support from regularization techniques are effective tools for modeling energy billing and revenue within academic institutions. ATBU's data shows reliable predictive relationships that can inform practical decision-making.

**Recommendations:**

- Institutional adoption of regularized regression models to improve forecasting.
- Expand the dataset across multiple institutions to enhance model generalizability.
- Explore Bayesian estimation frameworks for uncertainty quantification (e.g., empirical Bayes, credible intervals) arXivWikipedia.

## REFERENCES

Aldarraji, M., Vega-Márquez, B., Pontes, B., Mahmood, B., & C. Riquelme, J. (2024). Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models.

Dai, X., Gakidou, E., & Lopez, A. D. (2022). Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. Tobacco control.

Dalitz, C. (2018). Construction of Confidence Intervals.

FatehiJananloo, M., Stopps, H., & J. McArthur, J. (2023). Exploring Artificial Intelligence Methods for Energy Prediction in Healthcare Facilities: An In-Depth Extended Systematic Review.

Gao, Y., Liu, W., Wang, H., Wang, X., Yan, Y., & Zhang, R. (2023). A review of distributed statistical inference.

Gelman, A. & Vehtari, A. (2020). What are the most important statistical ideas of the past 50 years?.

Hopf, K., Hartstang, H., & Staake, T. (2023). Meta-Regression Analysis of Errors in Short-Term Electricity Load Forecasting.

Khajeh, H. & Laaksonen, H. (2022). Applications of Probabilistic Forecasting in Smart Grids : A Review.

Peñalvo, F. J. G. & Ingelmo, A. V. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. IJIMAI.

Priesmann, J., Münch, J., Ridha, E., Spiegel, T., Reich, M., Adam, M., Nolting, L., & Praktiknjo, A. (2021). Artificial Intelligence and Design of Experiments for Assessing Security of Electricity Supply: A Review and Strategic Outlook.

R. Falconer, J., Frank, E., L. L. Polaschek, D., & Joshi, C. (2021). Methods for Eliciting Informative Prior Distributions: A Critical Review.

Rajwar, K., Deep, K., & Das, S. (2023). An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. Artificial Intelligence Review.

W. Champ, C. & V. Sills, A. (2022). Process, Population, and Sample: the Researcher's Interest.