Energy Dynamics in the U.S.: Sector-Level Analysis and Forecasting with GDP and Population Factors

Background

Energy consumption is a critical component of economic activity and environmental sustainability. Understanding the patterns and drivers of energy consumption is essential for effective policy formulation and energy management. Over the past few decades, the United States has experienced significant changes in energy consumption across various sectors, influenced by economic growth, technological advancements, and demographic shifts.

The primary sectors consuming energy in the US are residential, commercial, and industrial. Each sector exhibits distinct consumption patterns influenced by factors such as seasonal weather variations, economic cycles, and policy changes. Analyzing these patterns can provide insights into the underlying dynamics of energy demand and identify opportunities for improving energy efficiency and sustainability.

Economic indicators such as Gross Domestic Product (GDP) and population growth are commonly used to understand and predict energy consumption trends. GDP, as a measure of economic activity, is closely linked to industrial and commercial energy demand, while population growth affects residential energy consumption. Exploring the relationships between these exogenous factors and energy consumption can enhance the accuracy of energy forecasts and inform strategies for managing energy resources.

This study aims to analyze the characteristics of US energy consumption from 1973 to 2022, focusing on the residential, commercial, and industrial sectors. By examining the trends, seasonality, and volatility of energy consumption and their relationship with GDP and population, this research seeks to provide a comprehensive understanding of the factors driving energy demand. The insights gained from this analysis will be valuable for policymakers, energy analysts, and economists in planning and forecasting future energy needs.

Introduction

We elected to analyze US energy consumption – sector-level characteristics and predictability – along with the relationships between energy consumption and exogenous factors like GDP and Population. The significance of this project lies in its potential to provide insights into the patterns of US energy consumption over the past few decades. By identifying sector-level characteristics; residential, commercial, and industrial sectors, we will be able to predict future energy consumption more accurately. Furthermore, this will guide the analysis exploring the relationships between energy consumption, GDP, and population.

This study aims to explore the properties of energy consumption in the US from 1973 to 2022 by addressing the following questions:

1. **Are there specific characteristics of the time series representing energy consumption**

over time that contribute most to the predictability of the time series?

- To answer this question, we will perform a time series analysis for each sector (residential, commercial, and industrial) to identify trends, seasonality, cyclicality, and volatility. We will utilize various time series models, including regression splines, SARIMA and ARMA-GARCH, to evaluate which characteristics enhance predictability. Accuracy measures like Mean Absolute Percentage Error (MAPE) and Precision Measure(PM), residual analysis, hypothesis testing, and visual representations of predictions will be used to assess model performance.

2. **Is there any relationship between the trends of energy consumption and GDP?**

- We will analyze the correlation between GDP and energy consumption trends by comparing the trend and seasonality components for each sector's time series with GDP. We will employ Vector Autoregressive (VAR) and Vector Autoregressive with Exogenous Variables (VARX) models to examine how changes in GDP influence energy consumption. Granger-causality tests will be used to determine if GDP can predict energy consumption in different sectors.

3. **Is there any exogenous factor that helps in predicting energy consumption?**

- In addition to GDP, we will investigate the impact of population growth on energy consumption. Population data will be integrated as an exogenous variable in our models to evaluate its predictive power. We will perform a Granger-causality analysis to identify whether population changes can predict energy consumption trends. The predictive performance of models incorporating both GDP and population will be compared against models with only one exogenous factor.

According to the US Energy Information Administration (EIA), there are several key trends that we should be mindful of during our analysis [1]. There has been moderate growth of US energy consumption that grew slower than GDP during the last 20 years – explained by a shift to efficient energy sources [2][3]. We should also consider seasonal variations that impact energy consumption; residential heating demands increase during colder months, and air conditioning usage rises in the warmer months, affecting both residential and commercial. We should be aware of renewable energy trends and the policies around renewable energy adoption which can greatly impact energy consumption, especially in the industrial sector. These could be exogenous factors to support our prediction of future energy consumption.

Data

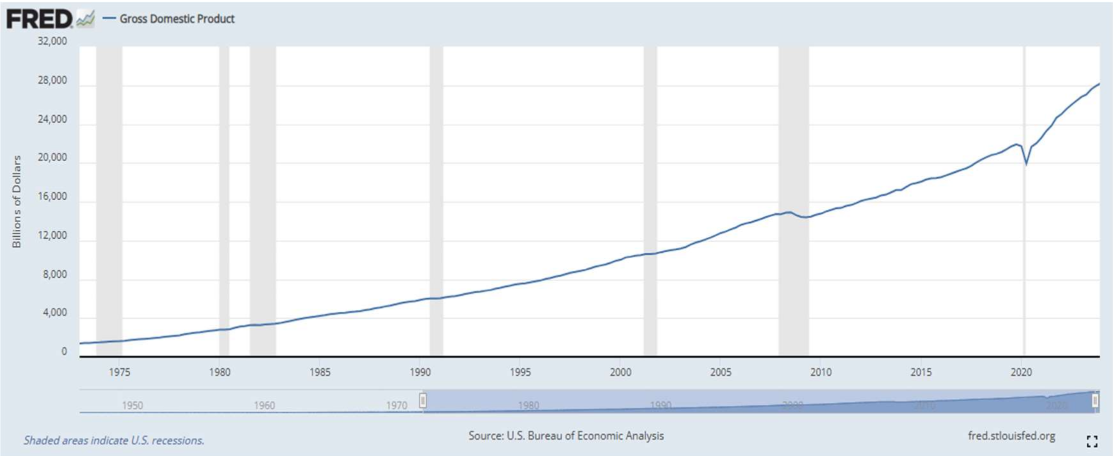
The energy consumption data was obtained from the US EIA [4] and contains the monthly energy consumption by sector, from January 1973 to April 2023. The monthly consumption data is divided into three sectors – residential, commercial, and industrial. Each sector is further broken down to provide a comprehensive description of the sector dynamics (Table 1).

Field Description	Type	Min	Max	Mean	Abbr.
Month (includes month and year)	chr	Jan-73	Dec-22		
Primary Energy Consumed by the Residential Sector	num	192	1488	585	Primary_RS

Electricity Sales to Ultimate Customers in the Residential Sector	num	136	570	309	ElecSales_RS
End Use Energy Consumed by the Residential Sector	num	491	1772	894	EndUse_RS
Residential Sector Electrical System Energy Losses	num	325	1145	663	Loss_RS
Total Energy Consumed by the Residential Sector	num	902	2808	1556	Total_RS
Primary Energy Consumed by the Commercial Sector	num	165	704	346	Primary_CS
Electricity Sales to Ultimate Customers in the Commercial Sector	num	114	461	283	ElecSales_CS
End Use Energy Consumed by the Commercial Sector	num	332	1096	629	EndUse_CS
Commercial Sector Electrical System Energy Losses	num	258	982	606	Loss_CS
Total Energy Consumed by the Commercial Sector	num	643	1873	1234	Total_CS
Primary Energy Consumed by the Industrial Sector	num	1466	2267	1807	Primary_IS
Electricity Sales to Ultimate Customers in the Industrial Sector	num	182	322	265	ElecSales_IS
End Use Energy Consumed by the Industrial Sector	num	1654	2467	2072	EndUse_IS
Industrial Sector Electrical System Energy Losses	num	405	777	576	Loss_IS
Total Energy Consumed by the Industrial Sector	num	2090	3082	2648	Total_I

[Table 1: Descriptive Analysis of the Energy Data]

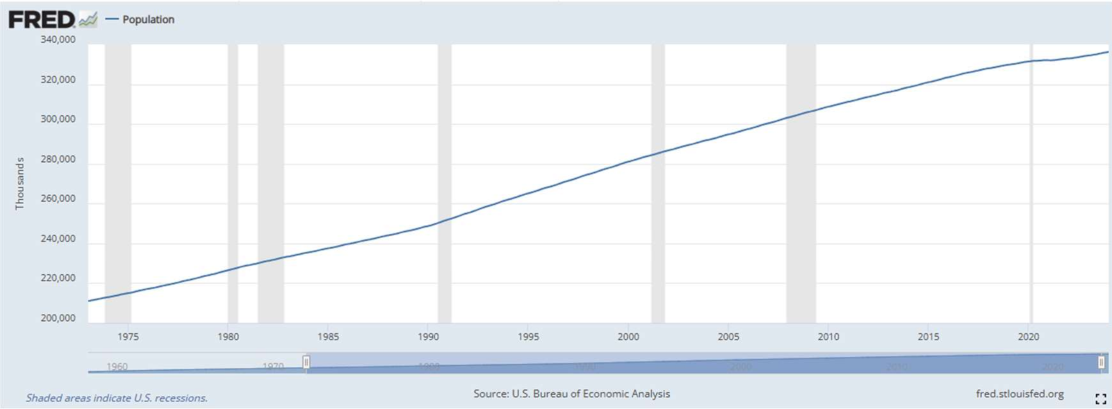
The GDP data, obtained from FRED Economic Data [5], serves as a proxy for economic activity and growth in our analysis. The dataset spans from 1973 to 2023 and includes quarterly GDP figures measured in billions of dollars. This data will enable us to investigate the relationship between economic growth and energy consumption across different energy sectors over time.



[Figure 1: Gross Domestic Product of the United States (1973-2023)]

The plot above illustrates the steady growth of the US GDP over the past five decades. Notably, periods of economic recession are shaded, highlighting their potential impact on energy consumption trends.

In addition to GDP, population growth is explored as an exogenous factor that may influence energy consumption patterns. The population data, also sourced from FRED Economic Data [6], provides monthly population figures from 1974 to 2023. This data is crucial for understanding how demographic changes impact energy consumption in the residential, commercial, and industrial sectors.

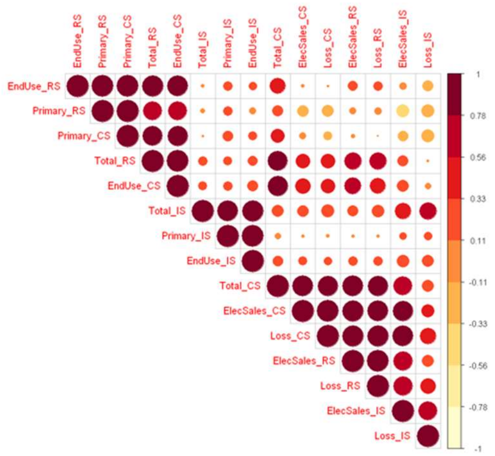


[Figure 2: Total Population of the United States (1975-2023)]

The population data visualization shows a consistent upward trend in the US population, reflecting continuous growth over the past several decades. This demographic trend is a key variable in analyzing energy consumption patterns.

Exploratory Data Analysis

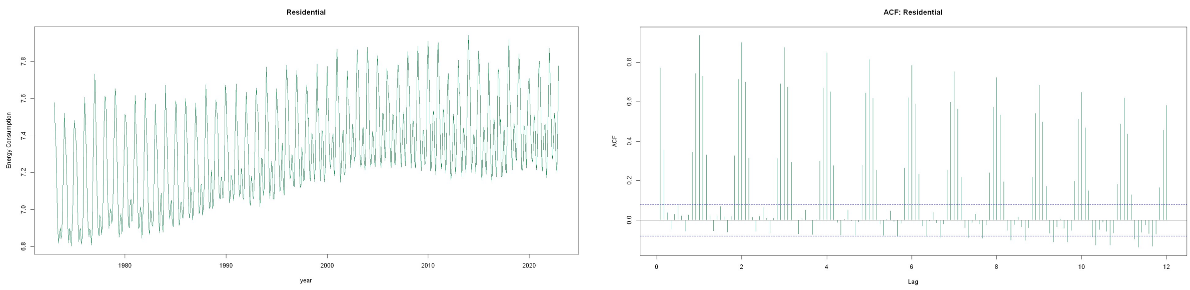
We performed exploratory data analysis to understand the variable distribution, correlations, outliers, trends, and seasonality. No discontinuities were found in the data. Histograms were created to examine the distributions, and boxplots were utilized to identify outliers. Outliers were detected in the Total Residential/Industrial, End Use Commercial/Industrial, Primary Industrial, and Loss Industrial variables. Given the context of our analysis, we chose to retain these outliers as they potentially represent natural variations due to unusual or extreme weather events, thus preserving valuable statistical insights and avoiding the underestimation of uncertainty [6][7].



In addition, a correlation matrix (Fig. A3) was used to identify highly correlated fields, which were expected as the total consumption variables are summations of different variables in the same sector. Thus, we will only focus on the total energy consumption per sector.

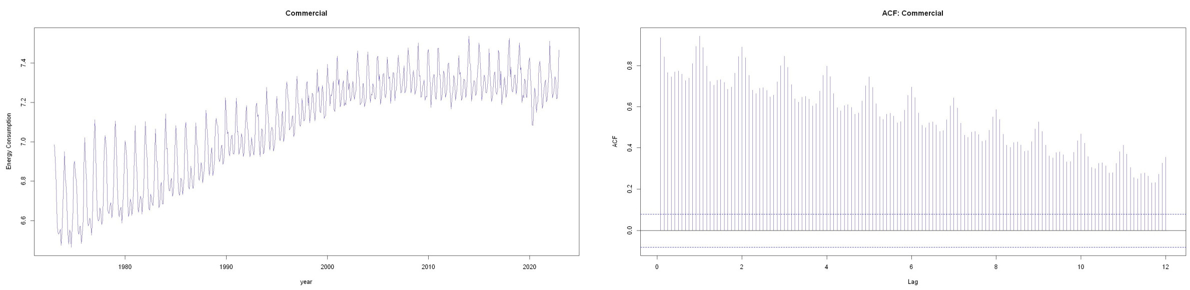
[Figure 3: Correlation Matrix]

After converting the total sector consumption variables into time series and plotting their respective time series and ACF plots, we observed distinct trends and seasonality patterns across the residential, commercial, and industrial energy sectors.



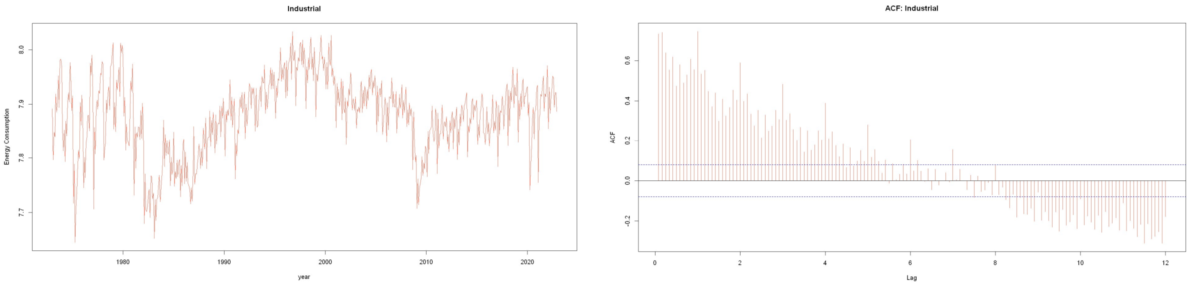
[Figure 4-1: Total Energy Consumption in *Residential* Sector]

For the **Residential sector** (Fig. 4-1), the time series plot demonstrates a clear seasonal pattern with peaks and troughs recurring annually. The ACF plot further confirms this seasonality, with significant autocorrelations at lags corresponding to 12 months, indicating strong yearly seasonal effects that need to be accounted for in the modeling process.



[Figure 4-2: Total Energy Consumption in *Commercial* Sector]

In the **Commercial sector** (Fig. 4-2), we observe a similar seasonal pattern in the time series plot, along with a noticeable upward trend over time. The ACF plot for the commercial sector also shows significant autocorrelations, emphasizing the importance of incorporating both trend and seasonality in the models.



[Figure 4-3: Total Energy Consumption in *Industrial* Sector]

The **Industrial sector** (Fig. 4-3) presents a more complex pattern. While there is evident seasonality and a general trend, the time series plot also indicates periods of higher volatility, particularly around economic events and recessions. The ACF plot shows significant autocorrelations, but with a pattern that suggests the presence of heteroskedasticity (aka fluctuating variance over time). This complexity in the industrial sector implies that more advanced modeling techniques may be required to adequately capture the underlying dynamics, including addressing the heteroskedasticity observed.

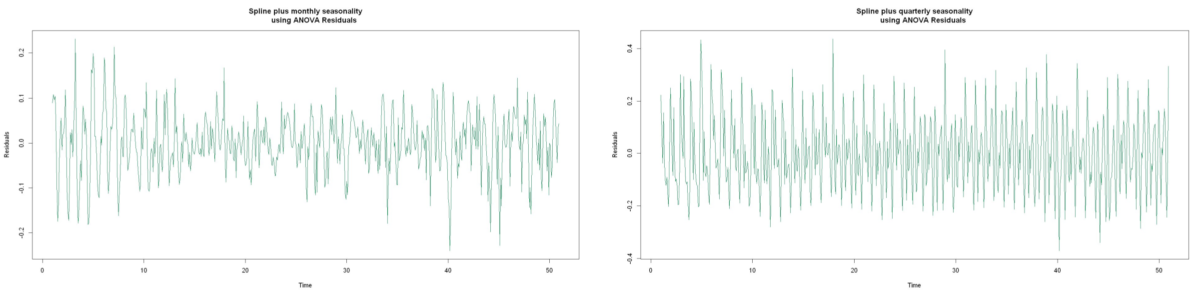
These observations led us to explore several modeling approaches to effectively account for the trend and seasonality in each sector. For the industrial sector, in particular, advanced techniques may be necessary to handle the observed heteroskedasticity and provide accurate forecasts.

Analysis of Time Series Models for Energy Consumption

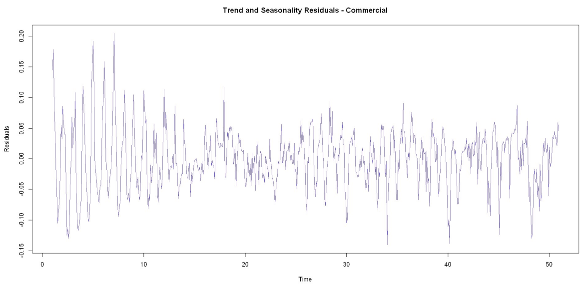
Understanding the dynamics of energy consumption in different sectors requires robust modeling approaches that can capture trends, seasonality, cyclicality, and volatility. This analysis aims to quantify the strength and significance of these time series characteristics in each sector—residential, commercial, and industrial. To achieve this, we explored several time series models to determine the most effective methods for predicting energy consumption and identifying the characteristics that enhance predictability in each sector.

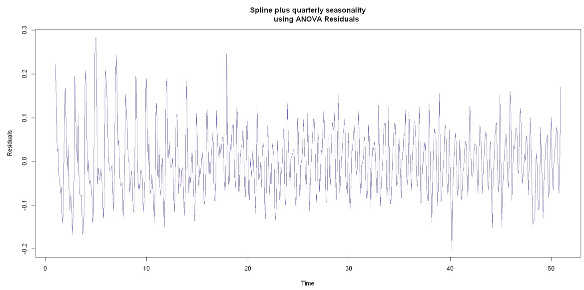
Regression Splines with Monthly and Quarterly Seasonality

The initial modeling using regression splines incorporated both monthly and quarterly seasonality via ANOVA. The residual plots (Figures 5-1, 5-2, 5-3) for the residential, commercial, and industrial sectors indicated that while these models captured some seasonality, the residuals still exhibited variability and patterns that suggested an incomplete model fit.

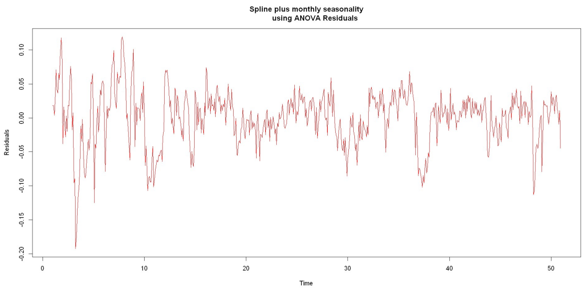


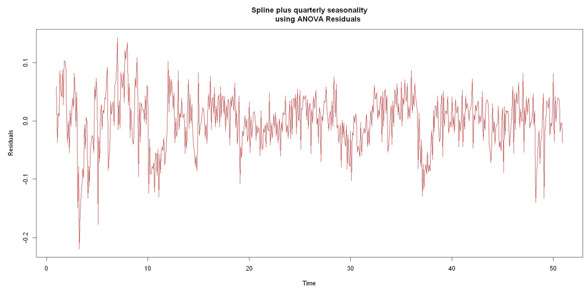
[Figure 5-1: Splines Residuals (*Residential*)]





[Figure 5-2: Splines Residuals (*Commercial*)]

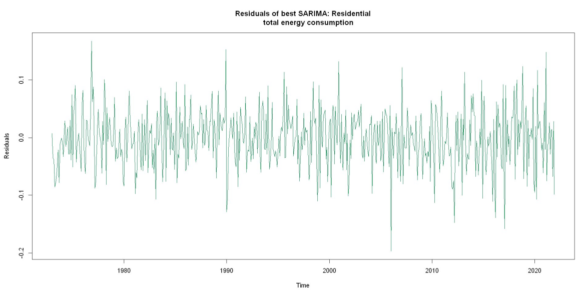


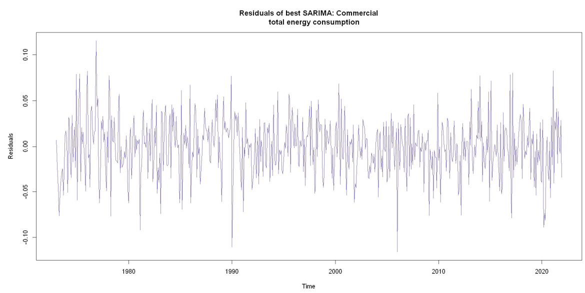


[Figure 5-3: Splines Residuals (*Industrial*)]

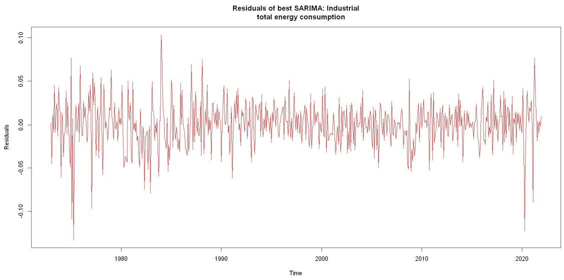
SARIMA

Next, we applied SARIMA models to each sector. Compared to regression splines, the SARIMA models proved to be more effective in capturing both trend and seasonality. This is evident from the residual plots for the SARIMA models (Figures 6-1, 6-2, 6-3), which show reduced variance and a better fit. Additionally, the SARIMA models exhibited superior accuracy measures (Table 2), indicating their enhanced predictive accuracy. The MAPE and PM were used to evaluate the prediction performance along with plotting the predictions and confidence intervals against the original data. The results were encouraging as the lags fell within the confidence intervals and the Precision Measures were relatively small (Table 2).





[Figure 6-1: SARIMA Residuals (Residential)] [Figure 6-2: SARIMA Residuals (Commercial)]

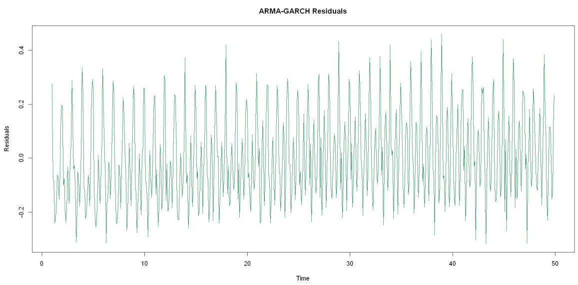


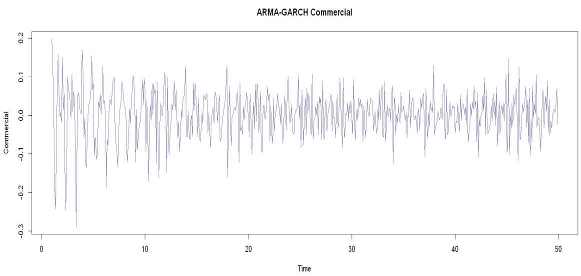
[Figure 6-3: SARIMA Residuals (Industrial)]

We optimized the orders of the SARIMA models to balance model fit and complexity using the AICc criterion. The optimal orders were as follows: Residential – (2,0,2)(1,1,1)[12]; Commercial – (2,0,1)(1,1,1)[12]; and Industrial – (2,0,2)(1,1,1)[12]. These orders suggest that the time series data for each sector has strong seasonal patterns with yearly cycles, and that the Residential and Industrial sectors may share more similar underlying dynamics compared to the Commercial sector.

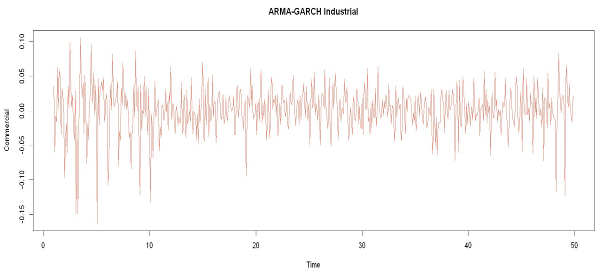
ARMA-GARCH

While ARMA-GARCH models captured volatility patterns, especially in the industrial sector, the SARIMA models generally provided a better fit for capturing both trend and seasonality in the residential and commercial sectors. This suggests that the residential and commercial time series data have more stable seasonal patterns and trends, making SARIMA a more appropriate model for these sectors, whereas the industrial sector's data likely exhibits higher volatility, which is better modeled by ARMA-GARCH.





[Figure 7-1: ARMA-GARCH Residuals (Residential)] [Figure 7-2: ARMA-GARCH Residuals (Commercial)]



[Figure 7-3] ARMA-GARCH Residuals (Industrial)

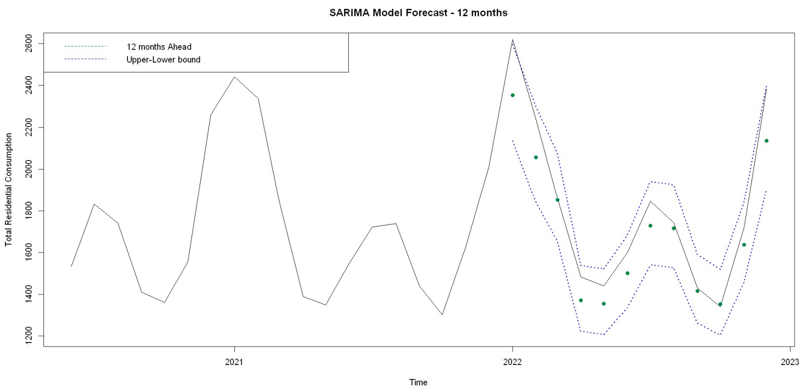
Are there specific characteristics of the time series representing Energy Consumption over time that contribute most to the predictability of the time series?

Having modeled the time series across different sectors using various approaches, we can now better understand the characteristics that influence their predictability.

To better understand what contributes to the predictability of energy consumption time series across different sectors, we can identify key characteristics by examining accuracy measures, residual analysis, hypothesis testing, and the visual representations of predictions. Let's explore these characteristics in detail.

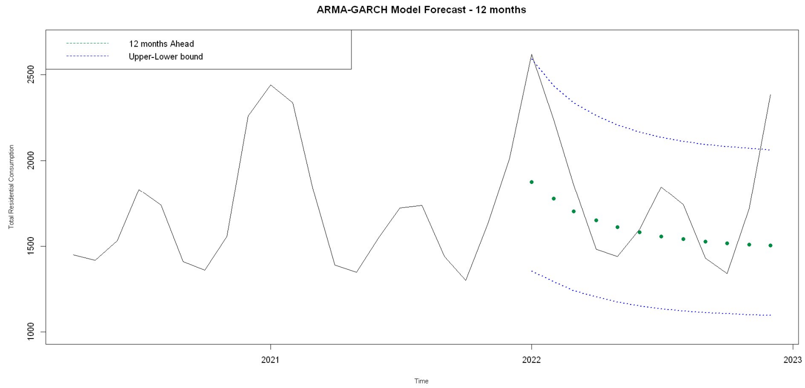
Residential Sector Findings

The SARIMA model for the residential sector shows no major serial correlations in the residuals, as indicated by the Box-Ljung test results (table 2). This suggests that the model effectively captures the underlying trend and seasonality in the data. Visually, the 12-month forecast of residential energy consumption using the SARIMA model aligns closely with the actual values, demonstrating its robustness in prediction (Fig. 8).



[Figure 8: SARIMA Forecast (Residential)]

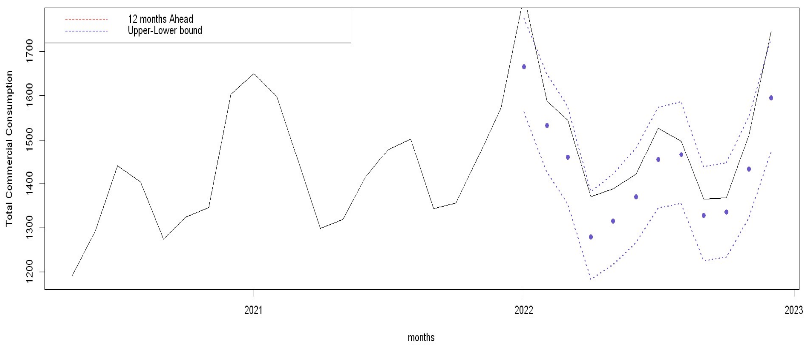
In contrast, the ARMA-GARCH model, while capturing some trends in the early predictions, struggles with predicting the inherent volatility in the data (Fig. 9). The MAPE for the ARMA-GARCH model is higher compared to the SARIMA model, indicating less accurate predictions (Table 2). The PM is close to 1, suggesting that the variability in the predictions mirrors the variability in the observed data over the prediction period (Table 2).



[Figure 9: ARMA-GARCH Forecast (Residential)]

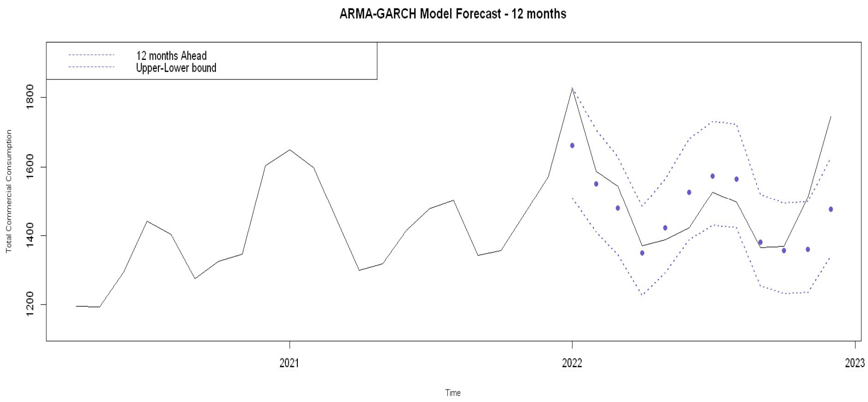
Commercial Sector Findings

Visually, the SARIMA model predictions for the commercial sector are very close to the actual observations for the 12-month forecast, indicating that both seasonality and trend are crucial for effective prediction in this sector (Fig. 10). The application of SARIMA to the commercial sector reduces variation in the residuals plot, although we still observe shocks of variation around 1990, 2008, and prior to 1980 (Fig. 6-2). This suggests the potential need for an ARMA-GARCH model to account for additional volatility in the data.



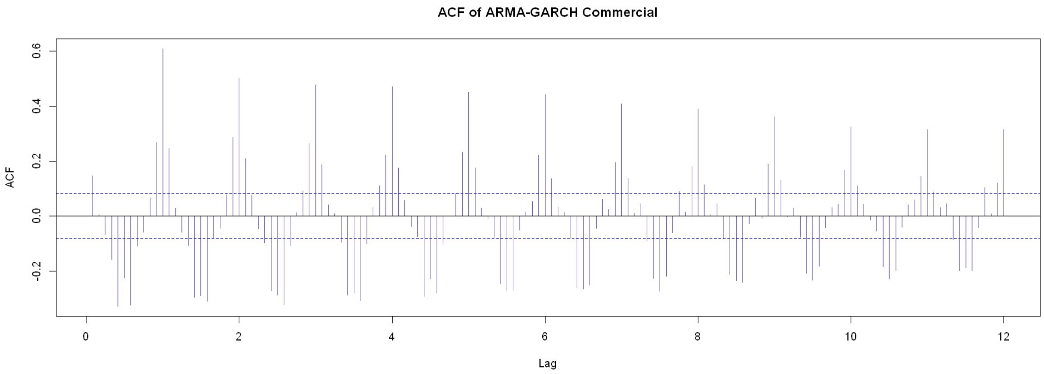
[Figure 10: SARIMA Forecast (Commercial)]

The ARMA-GARCH model forecast captures the predictions relatively closely to the actual observations, more effectively than in the residential sector (Fig. 11). This indicates a higher presence of volatility clustering in the commercial energy sector compared to the residential sector. The residuals from the ARMA-GARCH model still show shocks of variation during significant economic events, suggesting the influence of external factors like economic inflation, crashes, or geopolitical conflicts on energy consumption (Fig. 7-2).



[Figure 11: ARMA-GARCH Forecast (Commercial)]

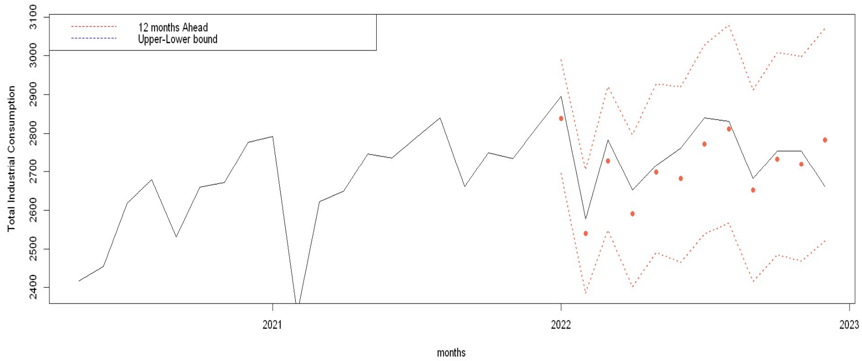
However, the ACF plot shows strong autocorrelation, and the seasonality is still visible(Fig. 12). Strong autocorrelation can indicate past values of the times series are significant predictors of future values. We still see extensive variation in the data, based on the ACF plots of both the residuals and the squared residuals, it does not appear to act as white noise, thus we should consider further analysis which can account for variation, which we will see in the Energy vs. GDP Analysis later.



[Figure 12: ACF of ARMA-GARCH (Commercial)]

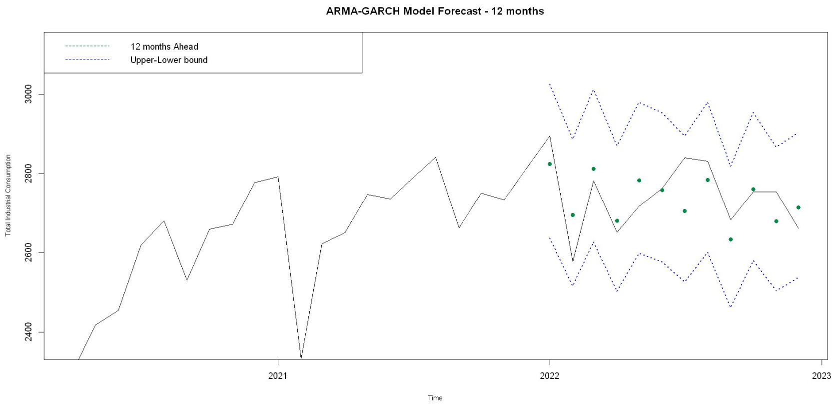
Industrial Sector Findings

The SARIMA model predictions for the industrial sector are visually close to the actual observations for the 12-month forecast, demonstrating that both seasonality and trend are essential for effective prediction in this sector (Fig. 13). Applying SARIMA reduces variation, although clusters of volatility remain around 1980-2000 and after 2020 (Fig. 6-3). These periods may require the ARMA-GARCH model to address additional volatility in the data.



[Figure 13: SARIMA Forecast (Industrial)]

The ARMA-GARCH model forecast also captures predictions closely to actual observations, similar to the commercial sector (Fig. 14). This indicates significant volatility clustering in the industrial energy sector, more so than in the residential sector.



[Figure 14: ARMA-GARCH Forecast (Industrial)]

Model (Energy Sector)	MAPE	PM	Box-Test Ljung (P-values)
SARIMA (Residential)	0.051	0.102	0.337
ARMA-GARCH (Residential)	0.148	0.990	2.2E-16
SARIMA (Commercial)	0.049	0.361	0.048
ARMA-GARCH (Commercial)	0.051	0.599	2.2E-16
SARIMA (Industrial)	0.018	0.446	0.378
ARMA-GARCH (Industrial)	0.021	0.631	6.06E-13

[Table 2: Accuracy Measures for SARIMA and ARMA-GARCH Models]

The analysis of energy consumption time series across residential, commercial, and industrial sectors reveals that strong seasonal patterns are a critical characteristic contributing to the predictability of these series. In both the residential and commercial sectors, the presence of regular, recurring seasonal trends plays a pivotal role in forecasting accuracy. SARIMA models, which are specifically designed to capture and model these seasonal patterns, outperform ARMA-GARCH models by effectively handling both trend and seasonality. This ability to integrate seasonal terms without requiring separate differencing for trend and seasonality simplifies the model and leads to more accurate predictions.

In contrast, the industrial sector exhibits a different characteristic—significant volatility clustering. This characteristic is more pronounced in the industrial sector than in the residential and commercial sectors. ARMA-GARCH models, which are well-suited for capturing periods of high volatility followed by periods of low volatility, are particularly valuable in modeling this sector. The GARCH component in these models helps in managing the unpredictability and sharp fluctuations seen in industrial energy consumption. However, despite the industrial sector's volatility, SARIMA models still generally provide better overall fit and predictive accuracy, suggesting that while volatility is important, seasonality remains a more dominant factor in driving predictability.

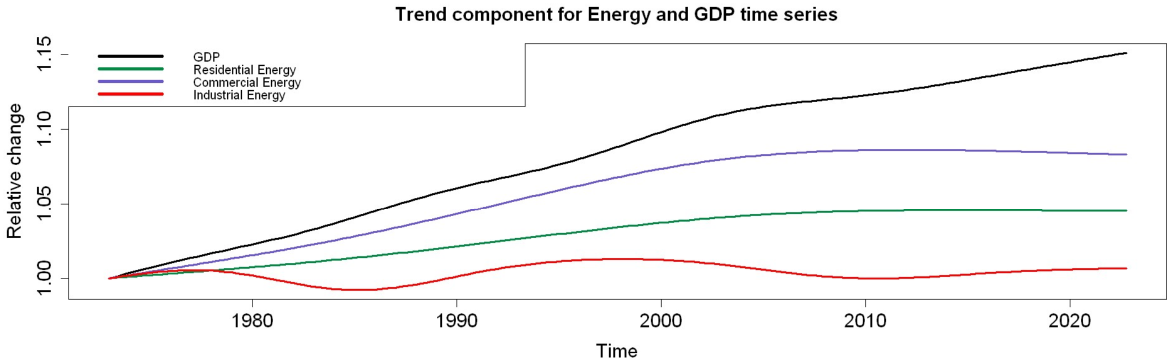
Overall, the findings indicate that seasonality is the most crucial characteristic for predicting energy consumption across all sectors, including industrial, where volatility is also significant. The superiority of SARIMA models in capturing seasonal trends across all sectors underscores the importance of this characteristic. Even in the industrial sector, where volatility clustering is evident, the predictability of the time series is still strongly tied to the seasonal component. This suggests that, across different energy sectors, a model's ability to effectively capture seasonality is key to enhancing the accuracy and reliability of energy consumption forecasts.

Is there any relationship between the trends of Energy Consumption and GDP?

To understand the relationship between energy consumption and GDP, we explored and compared the trend and seasonality components for various sector time series.

Trend Components

The trend components, obtained from GAMs, reveal a strong alignment between the trends in residential and commercial energy consumption with the GDP trend (Fig. 14). This close correlation suggests that fluctuations in GDP are closely mirrored by changes in energy consumption within these sectors. In contrast, the industrial energy consumption trend diverges significantly from the GDP trend, indicating that other factors beyond GDP may be driving energy usage in the industrial sector (Fig. 14).



[Figure 14: Trend Components for Energy Sectors and GDP]

Seasonality Components

The time series were detrended to isolate and examine seasonal effects, with quarterly seasonality coefficients estimated for each sector. Significant seasonal patterns were identified in the residential, commercial, and industrial energy sectors, indicating that energy consumption in these sectors is strongly influenced by seasonal factors. In contrast, no significant seasonal coefficients were found for GDP, suggesting that GDP does not exhibit a meaningful seasonal relationship with energy consumption. This implies that while energy consumption varies predictably with the seasons, GDP fluctuations are driven by other, non-seasonal factors.

Stationarity and Model Selection

To further investigate the relationship between the energy sectors and GDP, we prepared the data by differencing the time series with a lag of four quarters to achieve stationarity. Since GDP data is quarterly, we aggregated the energy consumption data accordingly.

We treated GDP as an endogenous variable, allowing us to capture the dynamic interrelationships between the variables over time. A Vector Autoregressive (VAR) model of order 1 was selected using the Bayesian Information Criterion (BIC).

To verify the assumptions of the VAR model, we conducted several tests. The small p-value from the *ARCH test* suggested heteroskedasticity, meaning the variance of the residuals is not constant over time. The small p-value from the *Portmanteau Test* indicated autocorrelation in the residuals, suggesting that past values have predictive power over current residuals. The *JB-Test* indicated a departure from normality in the residuals, meaning that the residuals do not follow a normal distribution. Finally, assumptions of stationarity were confirmed through root analysis, as all roots of the VAR model were inside the unit circle, indicating stability in the time series.

Granger-Causality Analysis

To evaluate whether GDP has a Granger-causality relationship with any of the energy consumption sectors, we applied the Wald test.

Variable Relationships	Wald Test (P-Value)
Residential – GDP	0.730
Commerical – GDP	0.036
Industrial – GDP	9E-06

[Table 3: VAR Models Wald Test – Granger-Causality of GDP as endogenous factor]

These results suggest that GDP has a Granger-causal relationship with the Industrial and Commercial sectors, meaning changes in GDP influence industrial and commercial energy consumption. No such relationship was found for the Residential sector. This finding aligns with the intuition that economic activity, represented by GDP, could drive industrial and commercial energy consumption.

Economic growth (GDP) can increase demand for industrial goods and services [8], leading to higher industrial activity and consequently, increased industrial energy consumption. When economic activity expands, industries tend to produce more goods, operate machinery at higher capacity, and engage in more construction and manufacturing activities, all of which require energy inputs within the industrial sector.

VARX Model with GDP as Exogenous Factor

We also created a VARX model with GDP as an exogenous factor and included a seasonal deterministic component with a quarterly cadence (season = 4). Exogenous variables influence endogenous variables but are not influenced by them.

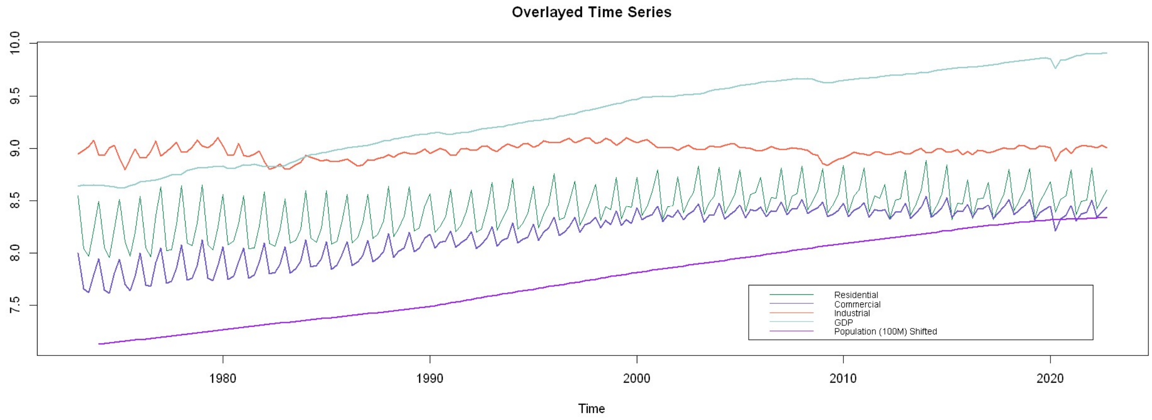
Variable Relationships	Wald Test (P-Value)
Residential – GDP	0.240
Commerical – GDP	0.0002
Industrial – GDP	0.0000

[Table 4: VARX Models Wald Test – Granger-Causality of GDP as exogenous factor]

We performed the Wald Test for Granger Causality of GDP as an exogeneous variable. These results suggest that changes in GDP have a measurable impact on the variables within the Commercial and Industrial sectors. For example, an increase in GDP may correspond to changes in variables such as industrial production, commercial activity, or business investment within these sectors, indicating that the Industrial and Commercial sectors are sensitive to changes in the broader economic environment captured by GDP.

Is there any exogeneous factor that helps in predicting Energy Consumption?

Previously, we explored treating GDP as an exogenous factor in the time series. Here we will explore how population can help predict energy consumption. The population data was sourced from World Bank – Data [9] and aggregated to the quarterly frequency to match the granularity of the GDP time series. All five time series were overlayed in a single plot (Fig. 15).



[Figure 15: Trend Components for Energy Sectors and GDP]

Seasonality Components

To focus on seasonal effects, we detrended the population data by applying a second-order differencing, making it stationary and suitable for further analysis. We modeled the seasonal components using a Generalized Additive Model (GAM), as this approach is effective in capturing non-linear seasonal patterns that could influence energy consumption.

VARX Model with GDP and Population as Exogenous Factor

Given the complexity of energy consumption patterns, we employed a VARX model with population and GDP as exogenous variables. This approach was chosen to account for the interdependencies between energy consumption in different sectors and the external influences of population growth and economic activity. The VARX model enabled us to generate forecasts for each sector, considering these exogenous factors, and we evaluated the accuracy of these forecasts using Mean Absolute Percentage Error (MAPE) and Precision Measure (PM).

The optimal order of the VARX model was determined using the BIC criterion, which indicated an order of 1 as most suitable. Finally, we conducted diagnostic tests on the VAR model to ensure its robustness, including tests for ARCH effects, serial correlation, and residual normality. These tests were essential to confirm that the model's assumptions were met and that the predictions were reliable. Additionally, we performed a Wald test to specifically assess the significance of population effects on energy consumption across sectors, justifying the inclusion of population as an important exogenous variable in the model (Table 5).

Granger-Causality Analysis

Interestingly, our analysis revealed that population Granger-Causes both commercial and industrial energy consumption, with a significance level set at 0.1; however, no such relationship was observed for residential energy consumption (Table 5) . In other words, a change in population has a measurable impact on the variables being modeled within Commercial and Industrial sectors.

Variable Relationships	Wald Test (P-Value)
Residential – GDP	0.140
Commerical – GDP	0.075
Industrial – GDP	0.069

[Table 5: VARX Models Wald Test – Granger-Causality of Population]

Predictive Performance Evaluation

While our primary focus centered on causality analysis within the VARX framework, we also evaluated the predictive performance by integrating GDP and population size as exogenous factors. This integration notably enhanced predictive accuracy, especially in the industrial sector. The Granger-causal relationship between GDP, population, and industrial energy consumption highlights the predictive power of these variables. In other words, when we consider GDP and Population as exogenous variables in the time series, it indicates that changes in these variables have a measurable impact on the Commercial and Industrial energy consumption.

Our VARX model achieved a remarkable MAPE score of 0.008 for industrial energy consumption, outperforming univariate models. However, the predictive performance for the residential and commercial sectors using the VARX framework was comparatively inferior, indicating a nuanced interplay between exogenous factors and energy consumption dynamics across different sectors (Table 6). Therefore, by considering exogenous factors like GDP and population, we can improve the accuracy of energy consumption predictions, especially in the industrial sector.

Model – Energy Sector (Exogenous Variable)	MAPE	PM
VARX – Residential (GDP)	0.112	1.373
VARX – Commercial (GDP)	0.053	1.293
VARX – Industrial (GDP)	0.012	2.691
VARX – Residential (GDP & Population)	0.120	1.288
VARX – Commercial (GDP & Population)	0.074	1.593
VARX – Industrial (GDP & Population)	0.008	1.815

[Table 6: Accuracy Measures Across Various VARX Models and Sectors]

Conclusion

This project investigated the characteristics of energy consumption time series data across three sectors: Residential, Commercial, and Industrial. We examined the relationship between energy consumption trends in these sectors and GDP to identify key factors influencing energy consumption predictability and explore potential exogenous factors for forecasting.

Our analysis revealed distinct characteristics in the energy data, including a clear upward trend indicating increasing overall consumption and pronounced seasonality patterns reflecting predictable fluctuations throughout the year. In contrast, GDP data exhibited a steady and consistent upward trend, suggesting a growing economy.

To study energy consumption, we employed various methods, including spline models with monthly and quarterly seasonality, SARIMA, and ARMA-GARCH models. SARIMA models provided the most accurate forecasts, as evaluated using MAPE and the PM. Both metrics indicated a good fit with low error and minimal bias.

To explore the connection between energy consumption and GDP, we utilized Vector Autoregression (VAR) and Vector Autoregression with Exogenous Variables (VARX) models, treating GDP as both an endogenous and exogenous variable. Our results suggested a plausible relationship between GDP and energy consumption in the Industrial and Commercial sectors in both the VAR and VARX models. However, no significant connection was found between GDP and the Residential sector, implying that economic growth primarily impacts industrial and commercial energy use, likely due to increased production activity.

By incorporating population data, we demonstrated that population serves as a Granger-cause for both commercial and industrial energy consumption. Furthermore, the accuracy of the forecasts for industrial energy consumption significantly improved when both GDP and population were incorporated as exogenous factors in the model.

Future Considerations

To further enhance our understanding of energy consumption, future studies should consider incorporating additional economic indicators such as employment rates and consumer spending. These indicators could provide valuable insights into the complex relationship between economic activity and energy use. Additionally, while this study focused on the US market, conducting comparative studies across different countries would provide valuable insights and contribute to a broader understanding of the intricate dynamics between economic factors and energy consumption.

By refining our models and expanding the scope of our analysis, we can continue to improve the accuracy of energy consumption forecasts and develop a more comprehensive understanding of the factors driving energy use across different sectors.

References

[1] U. E. I. Administration, "Annual Energy Outlook 2020," https://www.eia.gov/aeo, Washington, 2020.

[2] "U.S. energy facts explained: U.S. total energy statistics," December 2022. [Online]. Available: https://www.eia.gov/energyexplained/us-energy-facts/data-and-statistics.php. [Accessed 16 March 2024].

[3] "CONSUMPTION & EFFICIENCY: Current Issues & Trends," December 2022. [Online]. Available: https://www3.eia.gov/consumption/issuestrends/. [Accessed March 2024].

[4] "Total Energy: Monthly Energy Review," February 2024. [Online]. Available: https://www.eia.gov/totalenergy/data/monthly/.

[5] "FRED Economic Data: Gross Domestic Product," 28 February 2024. [Online]. Available: https://fred.stlouisfed.org/series/GDP.

[6] G. . F. M. Nascimento, F. Wurtz, . P. Kuo-Peng, B. Delinchant, and . N. J. Batistela, "Outlier Detection in Buildings' Power Consumption Data Using Forecast Error," Energies, 2021.

[7] K. Ferguson, "When Should You Delete Outliers from a Data Set?," 6 3 2018. [Online]. Available: https://humansofdata.atlan.com/2018/03/when-delete-outliers-dataset/. [Accessed 15 03 2024].

[8] Yi Hu, Dongmei Guo, Mingxi Wang, Xi Zhang, and Shouyang Wang. "The Relationship between Energy Consumption and Economic Growth: Evidence from China's Industrial Sectors". August 2015. Energies 2015, https://doi.org/10.3390/en8099392.

[9] "FRED Economic Data: Population," 29 February 2024. [Online]. Available: https://fred.stlouisfed.org/series/POPTHM.