

STATS 531 midterm project

Abstract

This study examines monthly electricity load data from the Electric Reliability Council of Texas (ERCOT) from 2002–2025 to understand long-term demand trends and seasonal patterns. Hourly observations were aggregated to a monthly frequency to highlight dominant seasonal structure. Visual inspection and formal stationarity tests indicated that the series is non-stationary, motivating first-order and seasonal differencing.

Using an AIC model selection procedure, a $\text{SARIMA}(0,1,3) \times (2,1,2)_{12}$ model was identified as the best specification among candidate models. Parameter estimates suggest both short-term shocks and strong seasonal dependence. Residual diagnostics indicate that the model captures the primary dynamics of the series, with approximately mean-zero and stationary residuals, though minor short-lag dependence and heavy-tailed behavior remain.

Overall, the analysis reveals sustained growth and pronounced seasonality in ERCOT electricity demand, providing a useful framework for forecasting while highlighting the potential benefits of incorporating climate and demographic drivers in future work.

1 Introduction

Texas has experienced substantial population growth over the past several decades, making it one of the fastest-growing states in the United States. This growth has been accompanied by increased electricity consumption. The changing climate is also having an impact on the demand for electricity (Dessler 2024).

Given that Texas is a rapidly growing state being faced with changing climate conditions, understanding how demographic and climatic factors interact to shape electricity load is important for managing energy resources efficiently.

Insights gained from modeling electricity load can inform interested parties by improving demand forecasting, supporting climate adaptation strategies, and guiding investments in energy capacity and resilience of the power grid.

2 Data

2.1 Overview

The data used in this analysis is from the Electric Reliability Council of Texas (ERCOT) (Electric Reliability Council of Texas (ERCOT) 2026). The dataset consists of hourly electricity load measurements beginning in 1995 and covers multiple zones across the state of Texas. These data provide a comprehensive view of electricity demand patterns over time and across regions with differing climatic and demographic characteristics.

We restrict attention to observations from 2002 onward. This subset provides a perspective on long-term electricity load trends, reflecting the more recent population growth and climatic changes.

2.2 Aggregation

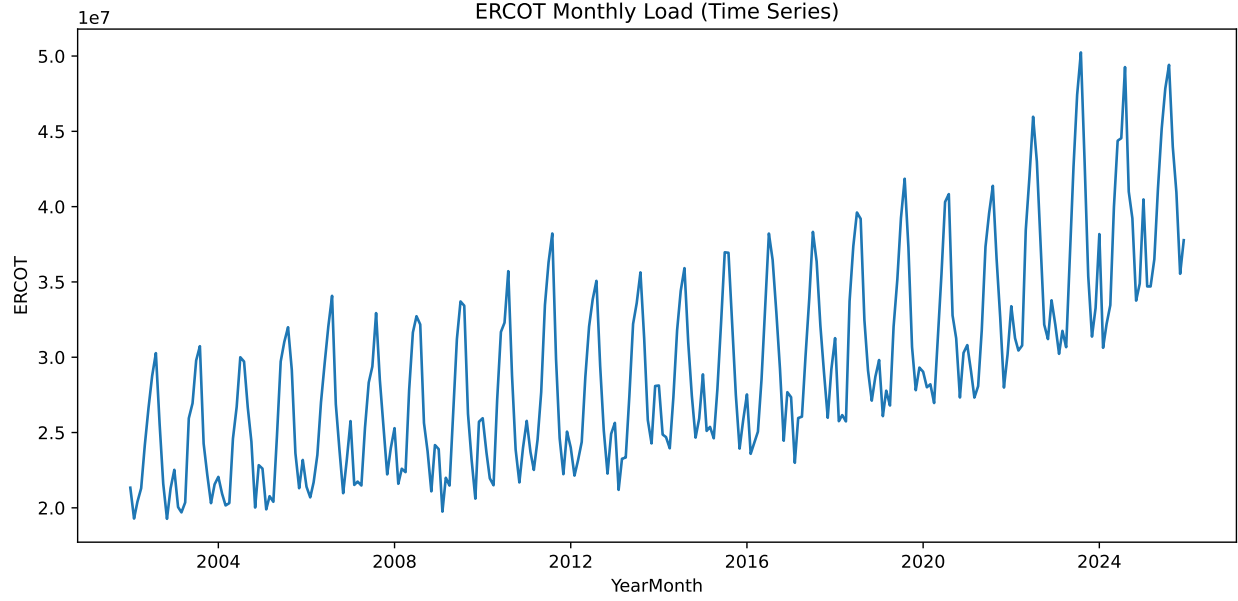
To examine seasonal effects more clearly, the hourly load data are aggregated to a monthly time scale. Monthly aggregation smooths high-frequency variability while preserving the dominant seasonal cycle, allowing for clearer visualization and modeling of recurring annual patterns and long-term trends. The data are shown below.

	YearMonth	ERCOT
0	2002-01-01	2.133632e+07
1	2002-02-01	1.928565e+07
2	2002-03-01	2.039984e+07
...
285	2025-10-01	4.094859e+07
286	2025-11-01	3.553874e+07
287	2025-12-01	3.777235e+07

3 Methodology

3.1 Visual inspection

First we plot our aggregated time series to inspect the trend and seasonality.



The series appears to be non-stationary from the plot. The mean increases with time, and the variance is all of equal scale. Visually we conclude that the series is covariance stationary, but not mean stationary. This motivates us to use more statistical tests to determine stationarity.

3.2 Stationarity tests

We now use the Augmented Dickey–Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to determine if the time series is stationary.

The ADF test tests the following hypotheses:

H_0 : The time series has a unit root (is nonstationary).

H_1 : The time series is stationary.

while KPSS tests the following hypotheses:

H_0 : The time series is stationary (level or trend stationary).

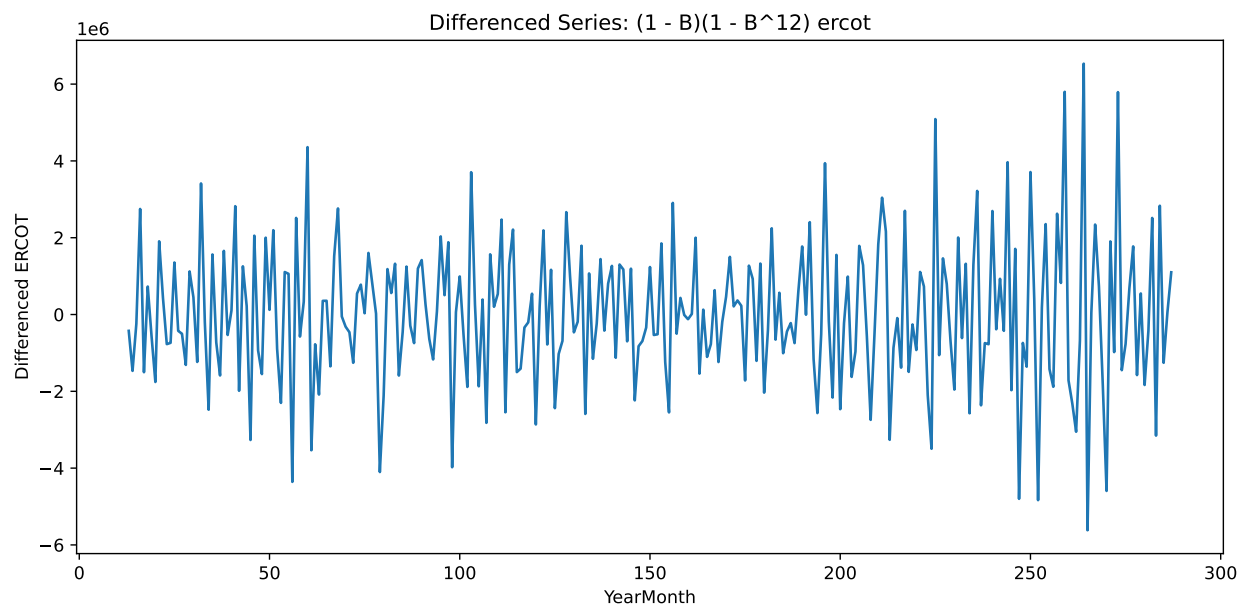
H_1 : The time series is nonstationary.

	test	series	stat	pvalue	lags	nobs
0	ADF	ERCOT	2.848886	1.00	12	275.0
1	KPSS(reg=c)	ERCOT	2.610717	0.01	9	NaN
2	KPSS(reg=ct)	ERCOT	0.264124	0.01	6	NaN

Our ADF test has a p-value of 1, and both KPSS tests have p-values of 0.01. The ADF test as well as both the level and trend KPSS tests indicate that the series is non-stationary and requires extra manipulation.

3.3 Differencing

For monthly data, a natural period for seasonal differencing is 12. Meanwhile, we also implement a non-seasonal first order differencing which makes the data more stationary while having minimal data loss.



We again apply ADF and KPSS tests on the differenced series.

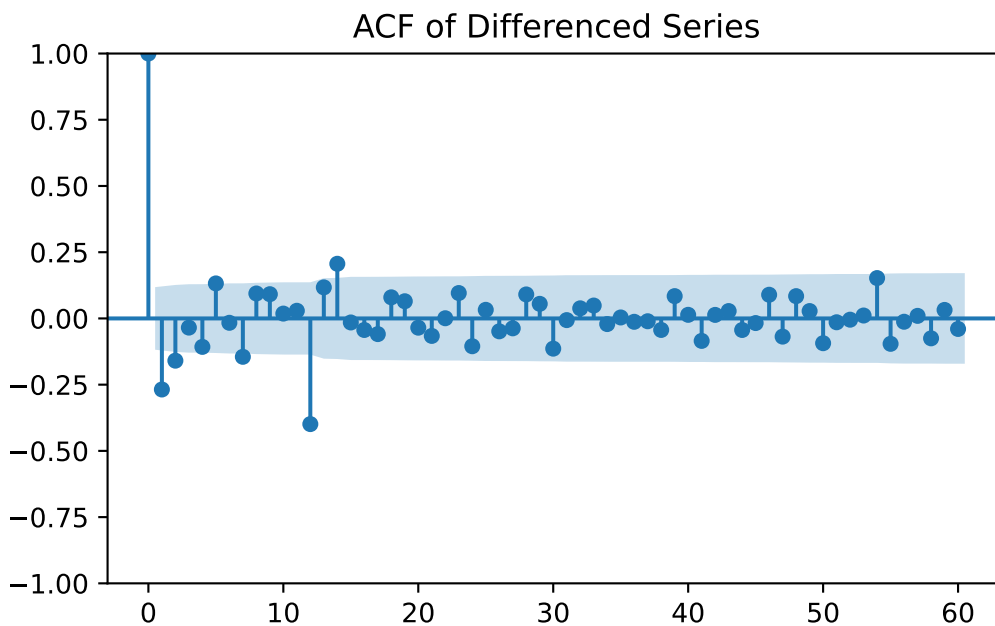
	test	series	stat	pvalue	lags	nobs
0	ADF	Differenced ERCOT	-7.774602	8.737033e-12	16	258.0
1	KPSS(reg=c)	Differenced ERCOT	0.150084	1.000000e-01	51	NaN
2	KPSS(reg=ct)	Differenced ERCOT	0.100715	1.000000e-01	51	NaN

The ADF test on the differenced series gives a p-value of $8.737(10^{-12})$, indicating stationarity. Both KPSS tests have p-values of 1, also suggesting that the differenced data is stationary.

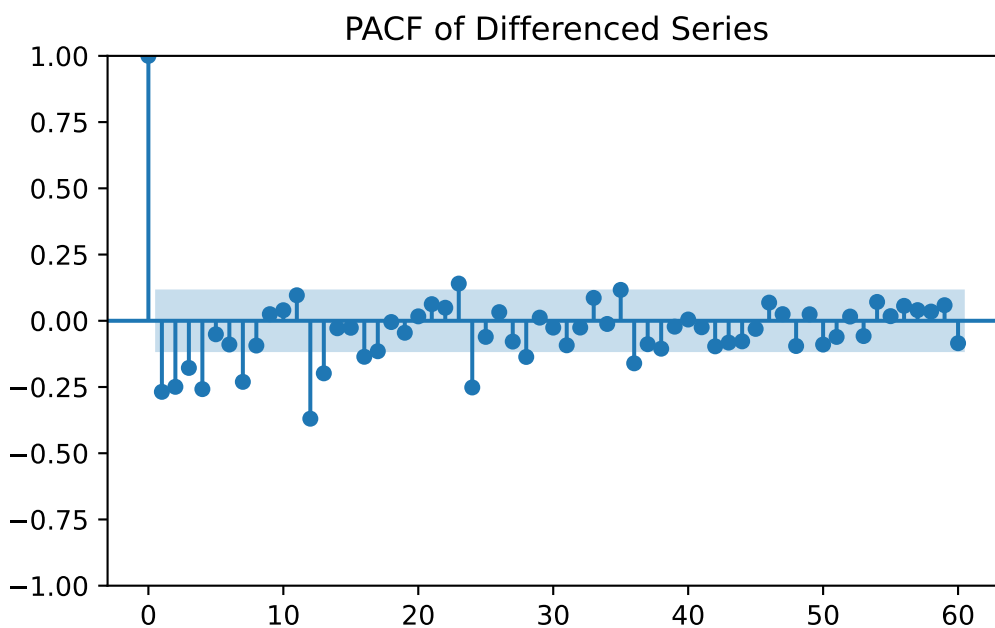
3.4 ACF and PACF

We further measure the stationarity of the differenced series using ACF and PACF plots.

<Figure size 3000x1200 with 0 Axes>



<Figure size 3000x1200 with 0 Axes>



The ACF of the differenced series exhibits rapid decay with most autocorrelations lying within the confidence bounds, indicating the absence of long-memory behavior. The PACF similarly shows only short-lag dependence, suggesting that the differencing strategy successfully achieved stationarity.

Residual dependence at lag 12 indicates the presence of stationary seasonal dynamics which justifies a SARIMA specification.

4 Results

4.1 AIC for SARIMA model specification

As suggested by the PACF plot, we choose to use a SARIMA model for our series. To select our parameters, we use Akaike information criterion (AIC), defined as

$$AIC = -2\ell(\hat{\theta}) + 2D$$

where $\ell(\hat{\theta})$ is the maximized log likelihood and D is the number of parameters. We will choose the set of parameters for our SARIMA model that minimizes this value.

	p	d	q	P	D	Q	AIC
0	0	1	3	2	1	2	7719.316566
1	2	1	3	2	1	2	7719.447080
2	1	1	3	2	1	2	7720.812957
...
141	2	1	0	0	1	0	8625.902708
142	1	1	0	0	1	0	8672.959460
143	0	1	0	0	1	0	8691.785154

4.2 Parameter estimates

Our AIC process chooses a SARIMA(0,1,3) \times (2,1,2)₁₂ model with an AIC value of 7719.317. So our model takes the form

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - B)(1 - B^{12})y_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)(1 + \Theta_1 B^{12} + \Theta_2 B^{24})\varepsilon_t$$

Dep. Variable:	ERCOT	No. Observations:	288
Model:	SARIMAX(0, 1, 3)x(2, 1, [1, 2], 12)	Log Likelihood	-3851.658
Date:	Wed, 18 Feb 2026	AIC	7719.317
Time:	23:54:38	BIC	7747.392
Sample:	0	HQIC	7730.620
	- 288		
Covariance Type:	opg		

coef	std err	z	P> z	[0.025	0.975]
------	---------	---	------	--------	--------

ma.L1	-0.5170	0.088	-5.878	0.000	-0.689	-0.345
ma.L2	-0.1706	0.110	-1.557	0.120	-0.385	0.044
ma.L3	-0.0867	0.092	-0.946	0.344	-0.266	0.093
ar.S.L12	0.0673	0.380	0.177	0.859	-0.677	0.811
ar.S.L24	-0.2876	0.126	-2.278	0.023	-0.535	-0.040
ma.S.L12	-0.6507	0.379	-1.719	0.086	-1.393	0.091
ma.S.L24	0.2513	0.187	1.342	0.179	-0.116	0.618
sigma2	2.752e+12	1.24e-14	2.21e+26	0.000	2.75e+12	2.75e+12

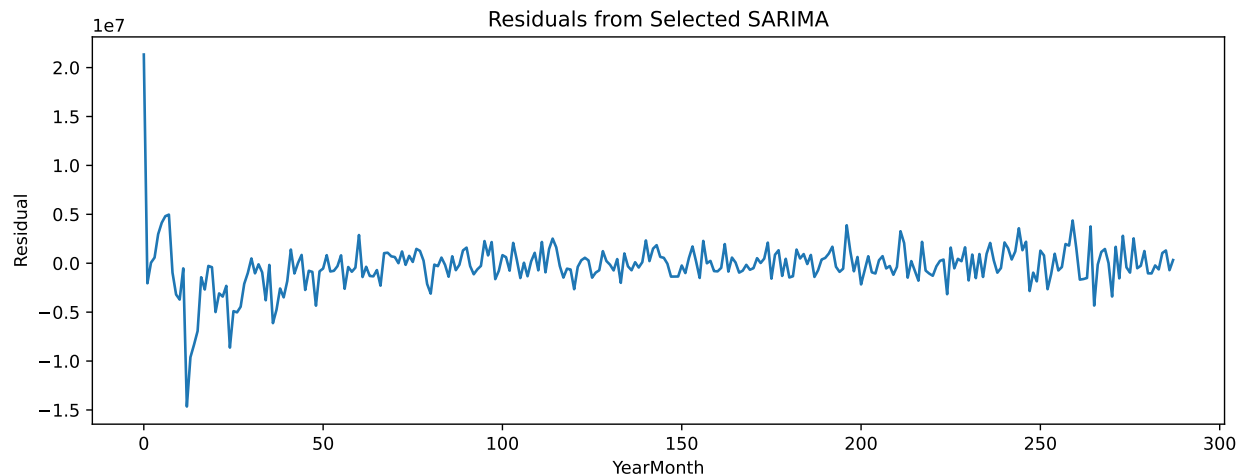
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	3.46
Prob(Q):	0.93	Prob(JB):	0.18
Heteroskedasticity (H):	1.54	Skew:	0.09
Prob(H) (two-sided):	0.05	Kurtosis:	3.55

The summary indicates the presence of both short-term and seasonal dynamics in ERCOT monthly electricity load. Parameter estimates reveal a significant nonseasonal MA component and a seasonal AR effect at lag 24, suggesting persistence of shocks across multiple seasonal cycles. Several additional parameters are statistically insignificant, including higher order MA terms.

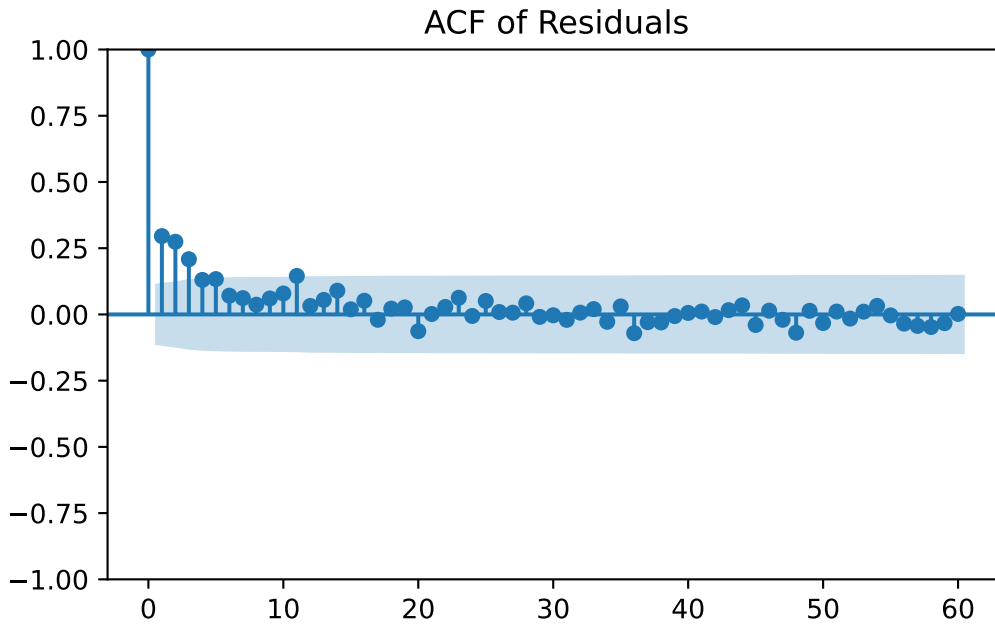
Overall, the model captures the dominant seasonal structure but may benefit from simplification due to some insignificant parameters. Some coefficient estimates should be interpreted cautiously due to potential near-singularity of the covariance matrix.

4.3 Residual diagnostics

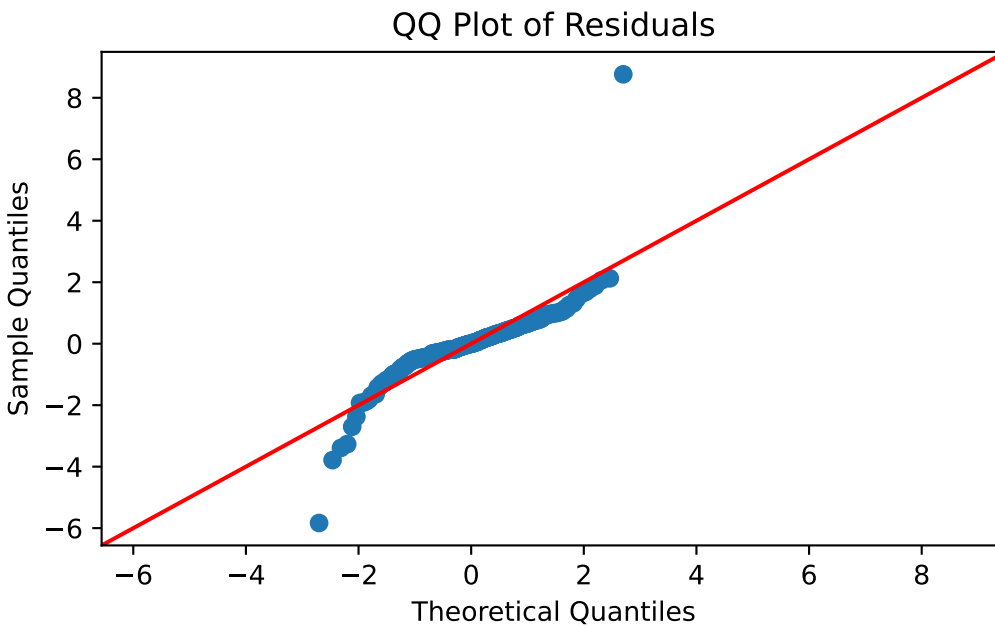
We now proceed to diagnose mean, independence, stationarity, and normality of the residuals.



<Figure size 3000x1200 with 0 Axes>



<Figure size 1800x1800 with 0 Axes>



The residuals from the selected SARIMA model fluctuate around zero with no obvious trend or remaining seasonality, suggesting the model captures the systematic structure of the series well, while the QQ plot suggests approximate normality in the central region with heavier tails. There is

some elevated volatility and a few large outliers in the sample, but the residual variance appears relatively stable over time.

The ACF of the residuals shows most autocorrelations lying within the 95% confidence bounds, with only small correlations at short lags, indicating little remaining dependence. The residual diagnostics are broadly consistent with white noise, supporting the adequacy of the fitted SARIMA model.

	ADF stat	p-value	lags	nobs
0	-3.873664	0.002237	16	271

The ADF test applied to the SARIMA residuals strongly rejects the null hypothesis of a unit root with $p\text{-value} = 0.00224$. This indicates that the residuals are stationary, providing further evidence that the fitted SARIMA model adequately captures the dynamics of the data.

5 Conclusion

In this study, we examined monthly ERCOT electricity load from 2002–2025 to understand long-term demand patterns and seasonal dynamics. Visual inspection and formal stationarity tests confirmed that the raw series exhibits non-stationary with an upward trend and strong annual seasonality. Applying first-order and seasonal differencing produced a stationary series suitable for modeling.

Using an AIC-based model selection procedure, we identified a $\text{SARIMA}(0,1,3) \times (2,1,2)_{12}$ model as the best specification. Parameter estimates suggest the presence of both short-term shocks and persistent seasonal dependence, including multi-year seasonal effects. Residual diagnostics indicate that the model captures the dominant structure of the data, with residuals centered around zero and approximately normal in the central region. Minor short-lag dependence and heavy-tailed behavior in the residuals suggest that some dynamics, particularly extreme demand fluctuations, remain unexplained.

These findings highlight a sustained increase in electricity demand alongside seasonal peaks, reflecting the combined influence of population growth and climate-driven energy needs. While the SARIMA framework provides a useful representation of ERCOT load dynamics, some limitations remain. The model assumes linear relationships and does not incorporate other variables such as temperature, demographic growth, or structural changes in energy consumption.

This analysis could be improved by incorporating climate variables within a SARIMAX framework, exploring nonlinear models, and addressing the heavy-tailed diagnostic problem. Overall, the SARIMA model offers a reasonable description of ERCOT monthly demand while providing a foundation for other modeling approaches that could include different perspectives on the research question.

Bibliography

- Dessler, Andrew E. 2024. "Climate Change Increases Energy Demand and Cost in Texas." <https://doi.org/10.22541/essoar.171561877.78686824/v1>.
- Electric Reliability Council of Texas (ERCOT). 2026. "Hourly Load Data Archives." Electric Reliability Council of Texas; https://www.ercot.com/gridinfo/load/load_hist. https://www.ercot.com/gridinfo/load/load_hist.