Predicting Average Electricity Prices from Data Center Investments and the AI Boom in the US

Conan Smallwood
Jasmine Kwok
Elizabeth Szeto
Erica Wang
Dominick Robinson

July 16, 2026

Executive Summary

In recent years, we have observed an AI boom in which public interest towards artificial intelligence rapidly increased, and investments towards data centers have exploded. At the same time, we have seen the prices of basic commodities increase, and in particular, the price of electricity has been going up continuously. With this, we asked the following questions: Do factors related to the AI boom (data center investments and AI search trends data) help predict electricity prices in the future? In either case, what model best predicts electricity prices?

The exploration incorporates three datasets. The first is the average electricity price in the U.S., which was obtained from the U.S. Bureau of Labor Statistics. Next was the investment data in data centers, which was taken from the U.S. Census Bureau. Finally, we acquired popularity scores of the topic AI on Google Trends as a proxy for public interest towards artificial intelligence. Upon obtaining the data we identified the overlapping time period between the three time series, linearly imputed missing electricity prices, and applied a log transformation and differenced the data to make each time series more stationary for subsequent modeling.

Once the data had been prepared, we fit and compared three models. First, we fit a seasonal ARIMA model to just the electricity prices data to represent a baseline model. If data center investments and AI trends data are indeed informative of electricity prices, we would expect a model containing these exogenous variables to outperform this baseline model. With this, we fit two time series regression model. In the first model we simply fit an ARIMA model to the electricity prices with the other two time series as exogenous regressors. In the second time series regression model, we fit the exogenous regressors at different lags to account for potential delayed effects. After fitting these models, we compared them using by looking at their RMSE and MAPE to identify the best model.

The results of the analyses were not what we had hoped. We were unable to observe any significant cross-correlation values for electricity prices with data center investments or Google AI trends data, meaning that these two time series are not informative of electricity prices. The model comparison results reflect this as well, where we observe the baseline SARIMA model having the lowest MAPE, with investment and trends data being non-significant regressors in the time series regression models. With this, we concluded that the factors relating to the AI boom that we looked at do not relate to or predict electricity prices in a meaningful way, and that the best model for forecasting prices is the baseline SARIMA model that solely uses past electricity prices. Next steps include conducting the same analysis on a shorter period of time to mitigate influences from evolving trends, incorporate a dataset that is a more accurate reflection of public interest towards AI, and identify other data sources that are more closely related to electricity prices for a more powerful time series forecast model.

Introduction

Over the past few years, the rise of artificial intelligence (AI) has led to a massive increase in data center construction across the United States. Large enterprises such as Google, Nvidia, and Amazon require more and more computing power to train AI models and run cloud-based services. To meet the growing demand, new data centers are being built at scale. These data centers consume a large amount of electricity, putting increasing pressure on power grids. This leads to an important question: as more data centers are built to support the use of AI, what is happening to electricity prices? Rising electricity prices have a direct impact on households, businesses, and the economy. As AI continues to grow rapidly, understanding how prices fluctuate and the magnitude of their change based on growing energy needs helps us to better prepare for the future.

This analysis aims to explore the relationship between data centers and electricity prices in the United States. We hypothesize that data center growth and AI-related activity influence US electricity prices. This leads to the research questions: **(1) do investments in data centers and growing public sentiment towards AI predict or relate to electricity prices, and (2) what model best predicts electricity prices?**

To answer these questions, we first conduct a time series analysis to investigate potential relationships using stationarity transformations and cross-correlation analysis. We then develop and compare several forecasting models, including a baseline seasonal ARIMA model and extensions incorporating exogenous and lagged predictors. Our results indicate that there is no strong evidence of a systematic relationship between AI-related activity, data center investment, and electricity prices, and that a univariate SARIMA model provides the most reliable forecasts.

Methods

Data

This analysis utilizes three monthly time series datasets: electricity prices, private investments, and Google search trends. The electricity price time series consists of 568 observations of monthly average electricity cost per kilowatt-hour in the United States, obtained from the U.S Bureau of Labor Statistics [4]. The series is not seasonally adjusted and spans November 1978 through February 2026.

The second dataset contains 397 private investments in the United States, sourced by the U.S Census Bureau. The data reported was the seasonally adjusted annual rate (SAAR) values in billions of dollars at a monthly frequency from January 2014 to January 2026 [3]. Although the original dataset contains multiple private investment categories, only the data centers investment series was used in this analysis.

The third dataset is Google Search Trends data for the category "AI" which is measured using a normalizing index ranging from 0 to 100 [1]. The data spans from January 2004 through March 2026 at a monthly frequency with a total of 267 rows.

To ensure comparability across datasets, all three time series were aligned to a common time window from January 2014 to January 2026. To stabilize variance and reduce heteroskedasticity, all series were transformed using the natural logarithm. Additional transformations were applied to address trend and seasonality, including first-order differencing and seasonal differencing at lag 12.

Missing Data

Missing values were minimal across the datasets. The electricity price series contained two missing observations, while no missing values were identified in the Google Trends or data center investment datasets. To preserve the continuity of the time series, the missing electricity price values were imputed using linear interpolation. This approach is appropriate given the smooth and gradually evolving nature of electricity prices. After imputation, a validation check confirmed that no missing values remained in the dataset.

Methodology

To address our research questions, we employed a two-stage time series analysis. First, we investigated whether AI-related activity and data center investment exhibit any meaningful relationship with electricity prices. Second, we developed and compared multiple forecasting models to determine the best predictive approach for electricity prices.

Relationship Analysis

To examine whether data center investment and AI-related activity are associated with electricity prices, we conducted a series of exploratory and diagnostic time series analyses. We first assessed the stationarity of each series using visual inspection, autocorrelation (ACF), and partial autocorrelation (PACF) plots.

To address nonstationarity, all series were transformed using logarithms and subsequently differenced. First-order differencing was applied to the Google Trends and data center investment series, while the electricity series required both first-order and seasonal differencing at lag 12 to remove trend and annual seasonality. The resulting differenced log-transformed series, which approximate stationarity, were then used for further analysis. Using these stationary series, we computed cross-correlation functions (CCF) between electricity prices and each predictor to identify potential lead-lag relationships.

Forecasting Models

Model 1: Baseline Seasonal ARIMA

As a baseline, electricity prices were modeled as a univariate time series using a seasonal ARIMA framework. The model was fit to the log-transformed electricity series with both first-order and seasonal differencing (lag 12) to account for trend and annual seasonality. Model orders were selected using the `auto.arima()` procedure with exhaustive search (`stepwise = FALSE`) to ensure optimal parameter selection. Model adequacy was evaluated using residual diagnostics, including residual time series plots, autocorrelation (ACF) and partial autocorrelation (PACF) of residuals, and normal Q-Q plots.

Model 2: ARIMA with Exogenous Regressors

To assess whether external variables improve predictive performance, we extended the baseline model by incorporating exogenous regressors, including a centered time trend, log-transformed Google Trends, and log-transformed data center investment. All regressors were centered to improve interpretability of the intercept and reduce potential multicollinearity. Forecasting with exogenous variables required generating future values of the predictors. To accomplish this, separate ARIMA models were fit to each predictor series, and their forecasts were used as inputs to the electricity forecasting model.

Model 3: Lagged Exogenous Model

To account for potential delayed effects, we constructed a third model using lagged versions of the predictor variables within a regression with ARIMA errors framework. Specifically, lagged values of the centered log-transformed Google Trends (lag 1) and data center investment (lag 3) were included alongside the centered time trend. We selected a one-month lag for Google Trends, assuming that shifts in public attention toward AI are reflected in energy demand after a short delay. Data center investment operate on a different timeline. Capital investment decisions take time to materialize into actual electricity consumption. A lag of 3 months reflects this longer pipeline between investment activity and the point at which that infrastructure is actually online and drawing power. These lagged variables were introduced to capture potential short-term delayed effects of AI-related activity and data center investment on electricity prices. Because lagging reduces the effective sample size, the electricity series was trimmed to match the aligned dataset used for model estimation. For forecasting, future values of the lagged predictors were constructed by carrying forward the most recent observed values, while the time trend was extended deterministically.

Model Comparison

All models were evaluated using standard time series diagnostic tools, including residual autocorrelation analysis and multiple goodness-of-fit measures such as RMSE(Root Mean Squared Error), MAE(Mean Absolute Error), and MAPE(Mean Absolute Percentage Error), along with visual inspection of forecast performance.

Results

Exploratory Data Analysis

Figure 1 displays the raw time series plots for all three variables over the shared window of January 2014 through January 2026. Electricity prices are relatively stable with clear seasonal oscillations from 2014 through early 2022, after which prices spike sharply, likely a consequence of the Russia-Ukraine war disrupting global energy markets, and continue to trend upward. A second notable sharp increase appears near the end of the series, potentially reflecting macroeconomic shocks such as the Trump administration's April 2025 tariff announcements. Google Search Trends for "AI" are nearly flat from 2014 through late 2022, at which point interest rises sharply, coinciding with the public release of ChatGPT 3.5 in late 2022 and the broader AI boom that followed. Data center investment, reported as a seasonally adjusted annual rate (SAAR) in billions of dollars, increases at a steady linear pace from 2014 through late 2022, after which the growth rate accelerates substantially, again aligning with the onset of large-scale AI adoption.

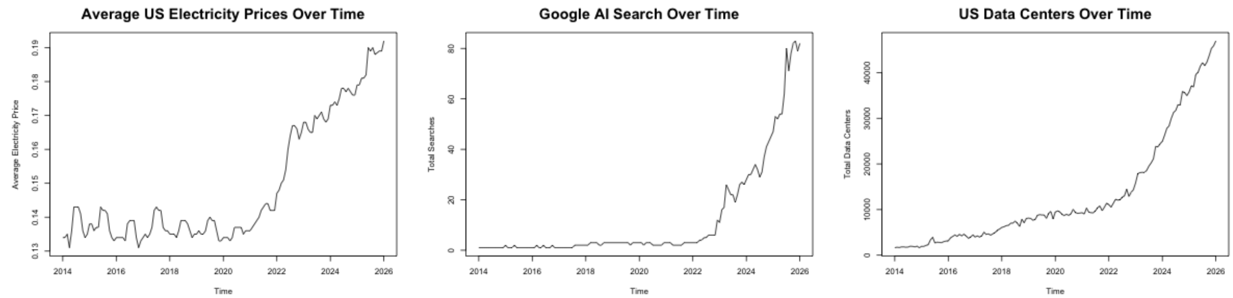


Figure 1: Raw time series plots for average U.S. electricity price (cents per kilowatt-hour), Google Search Trends index for “AI” and U.S. data center private investment (SAAR, billions of dollars), January 2014 to January 2026. Notable inflection points include the Russia-Ukraine war (early 2022) and the release of ChatGPT 3.5 (late 2022).

To stabilize variance, all three series were log-transformed prior to further analysis. The resulting log-transformed series are displayed in Figure 2. Electricity prices retain visible seasonal oscillations, and AI search trends and data center investment show persistent upward trends with accelerated growth after late 2022.

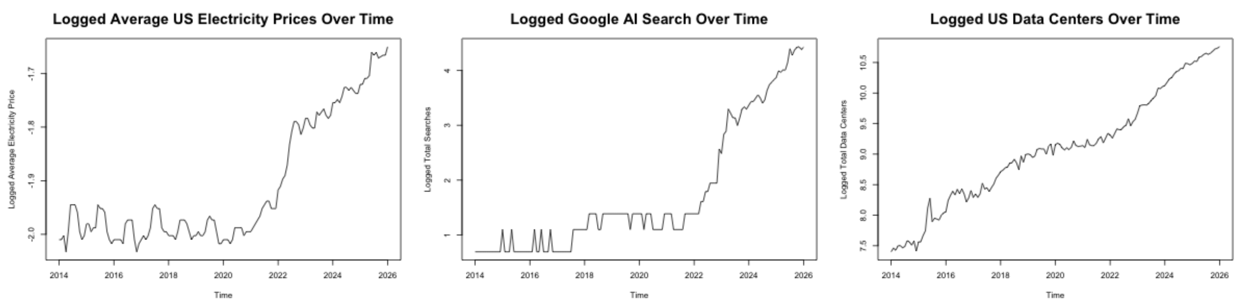


Figure 2: Log-transformed time series for all three variables over the common sample window, January 2014 to January 2026.

Seasonality

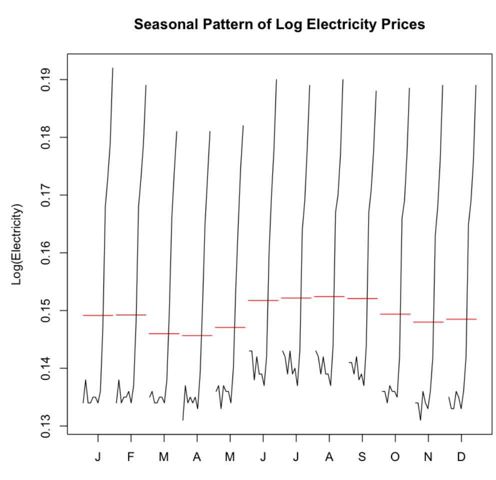


Figure 3: Month plot of log-transformed U.S. electricity prices. Each panel shows within-month variation across years. The red horizontal line marks the overall monthly mean.

A month plot of electricity prices confirms strong and consistent seasonal patterns. Average electricity prices per kilowatt-hour are highest in the summer months (June to August), driven by increased cooling demand, and exhibit secondary peaks in winter (January to February) due to heating demand. The spread within each month also widens over time, reflecting the overall upward trend in prices across the sample period. This persistent seasonal structure motivates the use of a seasonal ARIMA model for electricity prices.

Stationarity Assessment: ACF, PACF, and Differencing

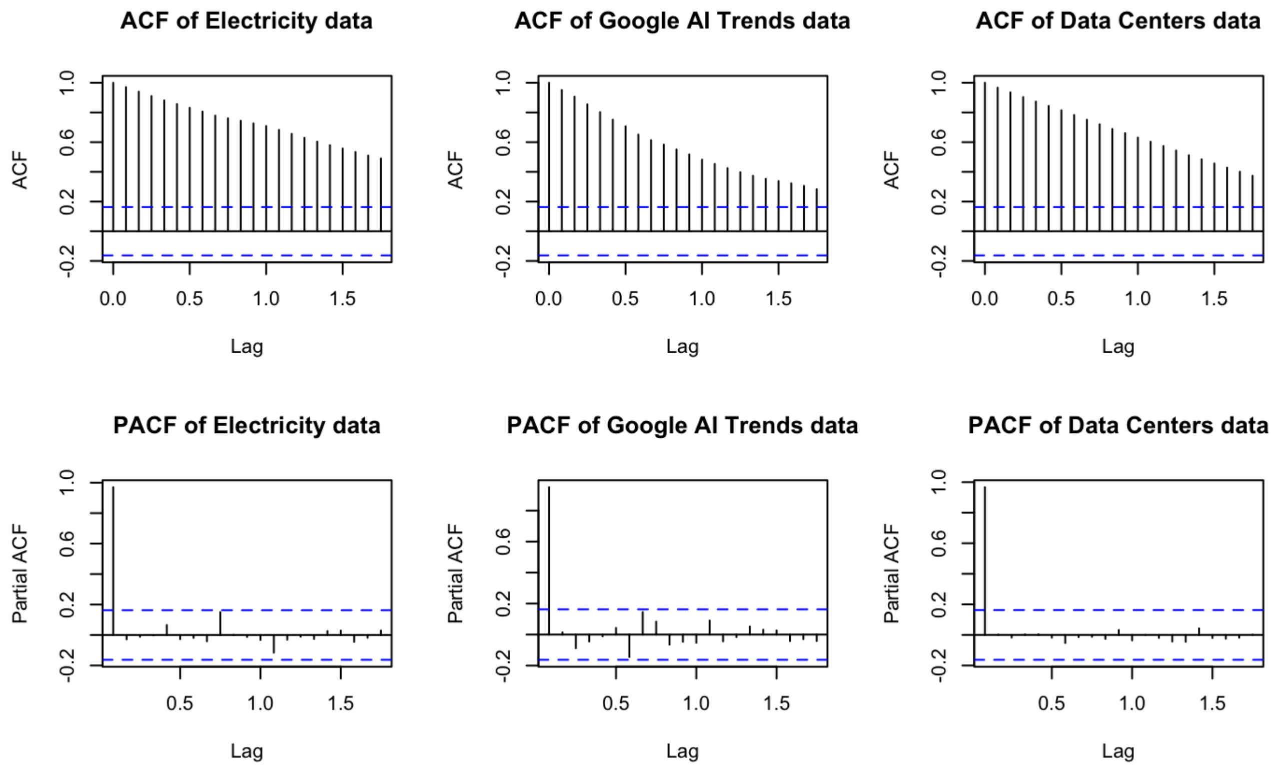


Figure 4: ACF and PACF for the raw series of electricity prices, Google AI search trends, and data center investment.

The ACF plots for all three series (Figure 4) exhibit slow, gradual decay rather than a sharp cutoff, indicating strong persistence and non-stationarity driven by underlying trends. The PACF for electricity prices shows a dominant spike at lag 1, suggesting a strong autoregressive component, while the absence of clear seasonal spikes in the AI trends and data center series indicates that their dependence is primarily trend-driven rather than seasonal.

To achieve stationarity, we apply a log transformation followed by first-order differencing (lag 1) to all three series, and additionally apply seasonal differencing (lag 12) to electricity prices to capture annual structure. While these transformations substantially reduce trend, the series are not fully stationary.

For electricity prices (Figure 5), the ACF and PACF of the doubly differenced series reveal both non-seasonal and seasonal dependence. At non-seasonal lags, the ACF decays gradually, indicating persistent dependence, while the PACF shows prominent spikes at the first few lags, particularly around lag 2, followed by smaller and less consistent spikes. Although the PACF does not exhibit a perfectly sharp cutoff, the dominant behavior at early lags suggests that a low-order autoregressive structure provides a reasonable approximation, with AR(2) chosen as a parsimonious specification. At seasonal lags, the ACF displays a strong and isolated spike at lag 12, with subsequent seasonal lags (24 and 36) falling within the confidence bounds, indicating a cutoff after one seasonal lag. The PACF shows several significant spikes at lags 2, 5, 8, 15, and 19, and does not exhibit a clear seasonal cutoff. Consequently, the seasonal specification is driven primarily by the behavior of the ACF at lag 12. Taken together, these observations suggest a SARIMA(2,1,0)(0,1,1)[12] model as an appropriate and parsimonious specification for the electricity price series.

For AI trends (Figure 6, left), residual autocorrelation persists at several early lags, and the series shows increasing variance after 2022, suggesting structural shifts associated with the rapid expansion of AI adoption that are not fully addressed by differencing. For data centers (Figure 6, right), a negative spike at lag 1 in the ACF points to mild overdifferencing, while the PACF indicates remaining short-term dependence.

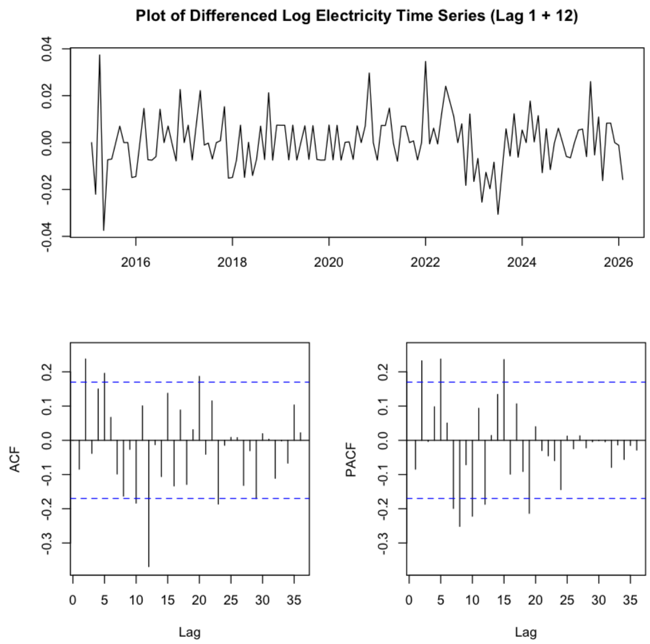
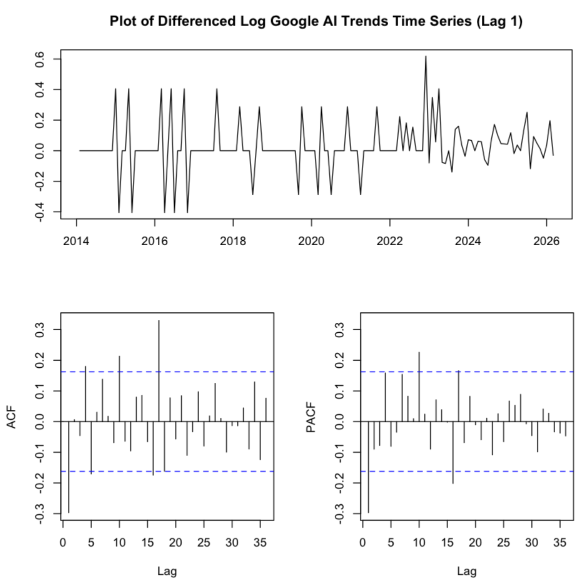
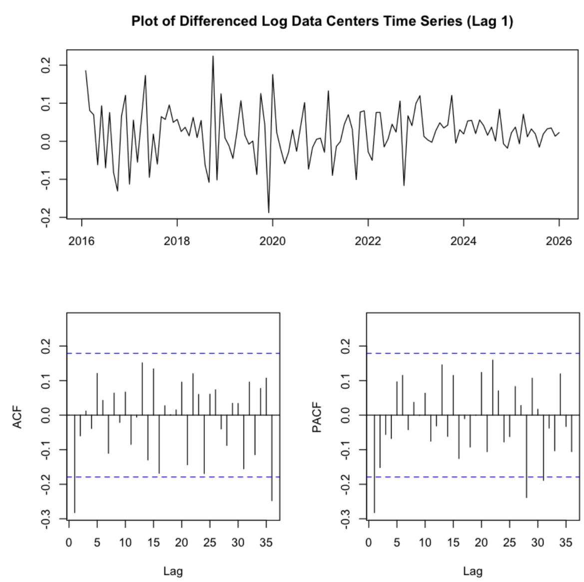


Figure 5: Time series plot, ACF, and PACF of the doubly-differenced log electricity price series.



(a) Differenced log Google AI search trends (lag 1)



(b) Differenced log data center investment (lag 1)

Figure 6: Time series plots, ACF, and PACF for the first-order differenced log Google AI search trends (left) and log data center investment (right). Both series exhibit reduced but not fully eliminated autocorrelation after differencing.

Cross-Correlation Analysis

Having assessed the stationarity properties of each series, we next examine whether any cross-serial relationships exist between the predictors and electricity prices. Cross-correlation functions (CCF) were computed between the differenced log electricity series and each of the two differenced predictor series. For both AI trends and data centers, most cross-correlations fall within the 95% confidence bounds across all tested lags, indicating no statistically significant lead–lag relationship between the predictor series and electricity prices. This suggests the cross-correlation between the differenced series provides little evidence that either predictor leads or lags electricity prices in a meaningful way. Because the stationarity assumptions are not fully satisfied, particularly for the AI trends and data center series, which exhibit strong persistent growth, these CCF results should be interpreted with caution. A VAR model was considered as an alternative multivariate framework that could jointly model all three series, but was ultimately not pursued given that the stationarity requirements for VAR were not sufficiently met, and the CCF analysis gave little indication of lagged cross-series dynamics.

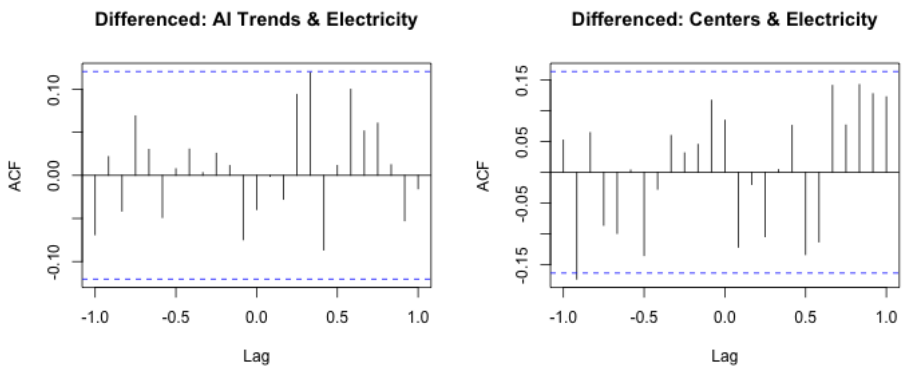


Figure 7: Cross-correlation functions (CCF) between the differenced log electricity price series and the differenced log Google AI search trends (left) and differenced log data center investment (right), computed at lags up to ±12 months.

Model Diagnostics

Baseline SARIMA$(2,1,0)(0,1,1)_{12}$ (Model 1)

The baseline model's seasonal MA term ($\hat{\theta}_1 = -0.627$) is large and statistically significant, indicating strong correction toward the previous year's seasonal level. Table 1 shows the estimated model coefficients and standard errors. The two autoregressive coefficients ($\hat{\phi}_1 = -0.047$, $\hat{\phi}_2 = 0.170$) are small relative to their standard errors and are not statistically significant, suggesting limited short-term autoregressive dependence beyond what is captured by the differencing and seasonal structure.

Seen in Figure 8, the "Observed and Fitted over Time" and "Observed vs. Fitted" plots show the model tracking the observed series closely, including the sharp price increase beginning in 2022. The scatter plot of observed versus fitted values yields a correlation of 0.997, showing an excellent fit. The residual time series plot shows no obvious systematic patterns, and the ACF and PACF of the residuals show nearly all lags within the 95% confidence bounds, indicating that little autocorrelation structure remains. The Q-Q plot of residuals follows the normal line closely through the central quantiles, with minor deviations in the tails, suggesting slight heavy-tailedness but no severe violation of normality. Overall, the baseline model fits well, with a training RMSE of 0.0094 and a MAPE of approximately 0.37%, indicating that the ARIMA structure alone accounts for most of the predictable variation in log electricity prices.

Figure 9 displays the 12-month ahead forecast from the baseline model. It indicates that Model 1 expects electricity prices to continue increasing over the next year. The solid black line extends the historical log electricity prices into the forecast horizon, while the dashed red lines represent the upper and lower bounds of the 95% prediction interval. Overall, the model projects a modest upward trend over the next 12 months, consistent with recent historical movement in the series. The relatively narrow prediction interval suggests high short-term confidence in the forecasts, which is likely driven by the strong seasonal and autoregressive structure captured by the model when fitted using only the electricity price data.

Table 1: Model 1 Coefficients: The significant seasonal MA term indicates that yearly patterns are the primary driver of price fluctuations.

Parameter	**Estimate**	**S.E.**	**t-statistic**
AR(1)	-0.0472	0.0859	-0.55
AR(2)	0.1696	0.0859	1.97
Seasonal MA(1)	-0.6270	0.0869	-7.21

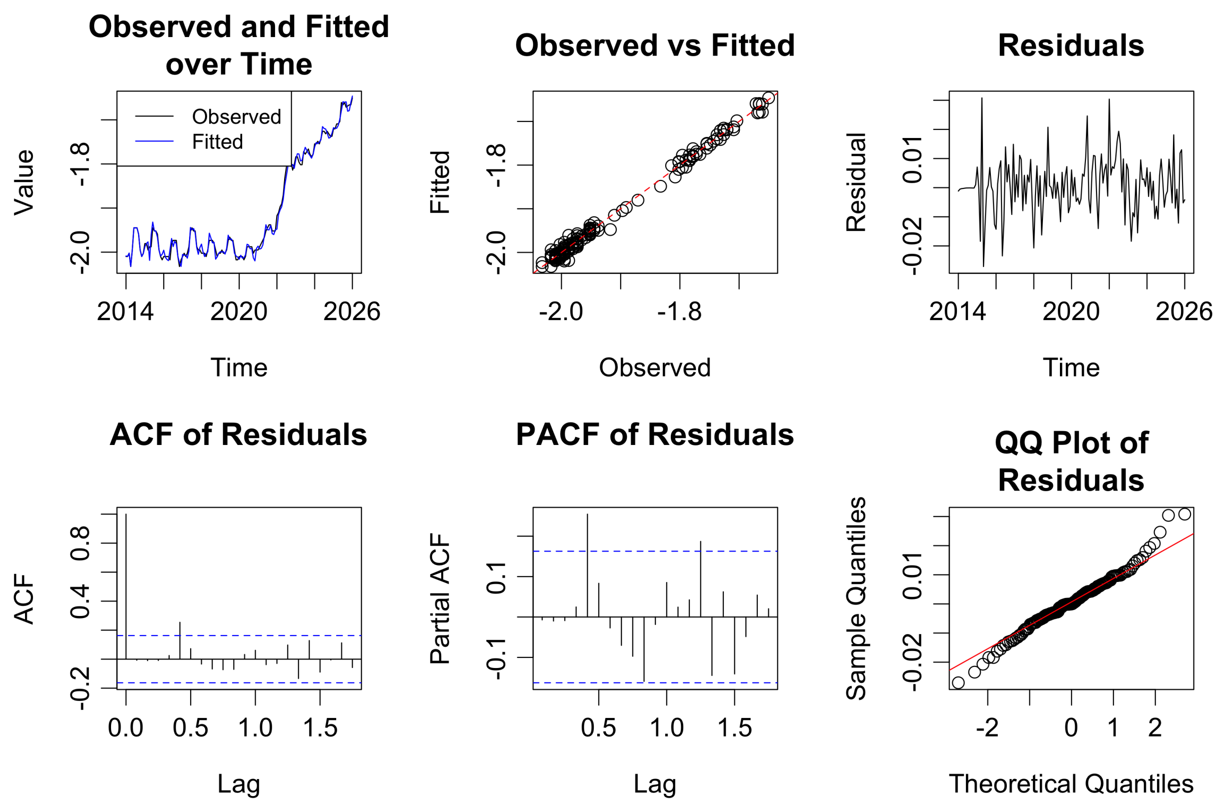


Figure 8: Baseline Model Diagnostics: A correlation of 0.997 and white noise residuals confirm the baseline SARIMA model captures almost all predictable variation.

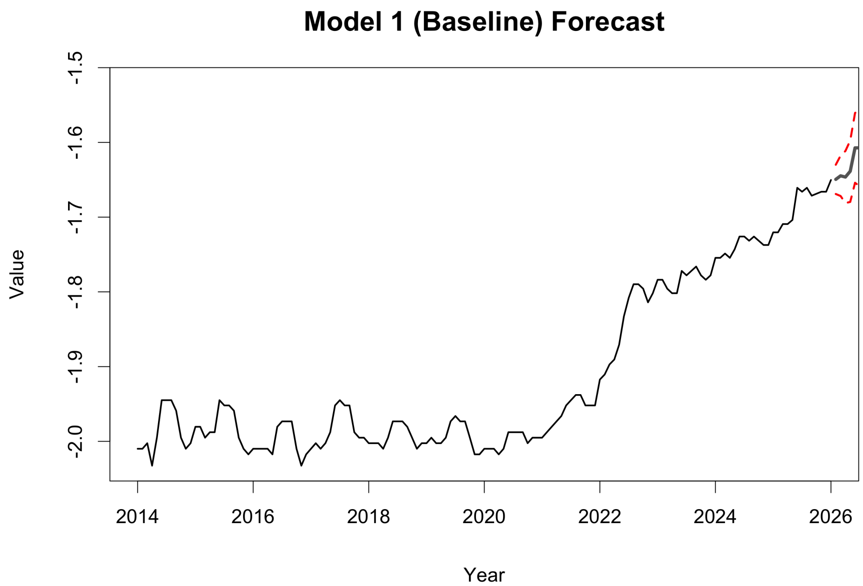


Figure 9: Baseline model 1 forecasts electricity prices likely increasing over next 12 months.

Model 2 Diagnostics

Model 2 extended the baseline by incorporating concurrent exogenous variables: a centered time trend, log-transformed Google AI Trends, and log-transformed data center investment. Upon including these predictors, the model structure transitioned to a regression with ARIMA$(1,0,0)(2,0,0)_{12}$ errors.

Table 2 shows the estimated model coefficients and standard errors. The AR(1) coefficient ($\hat{\phi}_1 = 0.971$) is close to 1, indicating that after conditioning on the external regressors, log electricity prices remain highly persistent from one month to the next. This level of persistence suggests the series is borderline non-stationary even after accounting for the trend regressor. The two seasonal AR terms ($\hat{\phi}_{s,1} = 0.441$, $\hat{\phi}_{s,2} = 0.374$) are both significant. Among the external regressors, only the linear time trend is statistically significant ($\hat{\beta}_{\text{trend}} = 0.029$, s.e. $= 0.013$); the coefficients for logged AI trends and logged data center investment are negligible relative to their standard errors.

Seen in Figure 10, the "Observed and Fitted over Time" and "Observed vs Fitted" plots (correlation = 0.996) confirm a strong in-sample fit. Residuals show no obvious structure in the ACF or PACF plots, and the Q-Q plot indicates approximately normal residuals with minor tail departures. Training RMSE is 0.0102 and MAPE is 0.394%.

Figure 11 shows that Model 2 similarly projects a continued upward trend in log electricity prices over the next 12 months. However, the prediction interval is slightly wider than that of the baseline, despite Model 2 incorporating additional exogenous information. This is because forecasting with exogenous regressors requires generating future values of those predictors via separate ARIMA models, and the uncertainty from those auxiliary forecasts propagates into the final electricity price forecast.

Table 2: Model 2 Coefficients: The time trend is significant, but concurrent AI and data center variables do not provide additional explanatory power.

Variable	Estimate ($\hat{\beta}$)	S.E.	t-statistic
Autoregressive (1)	0.9713	0.0151	64.32
Seasonal AR(1)	0.4406	0.0874	5.04
Seasonal AR(2)	0.3741	0.0922	4.06
Intercept	-1.8437	0.1018	-18.11
Linear Time Trend	0.0285	0.0125	2.28
Log AI Trends	-0.0001	0.0049	-0.02
Log Data Center Investment	0.0063	0.0096	0.65

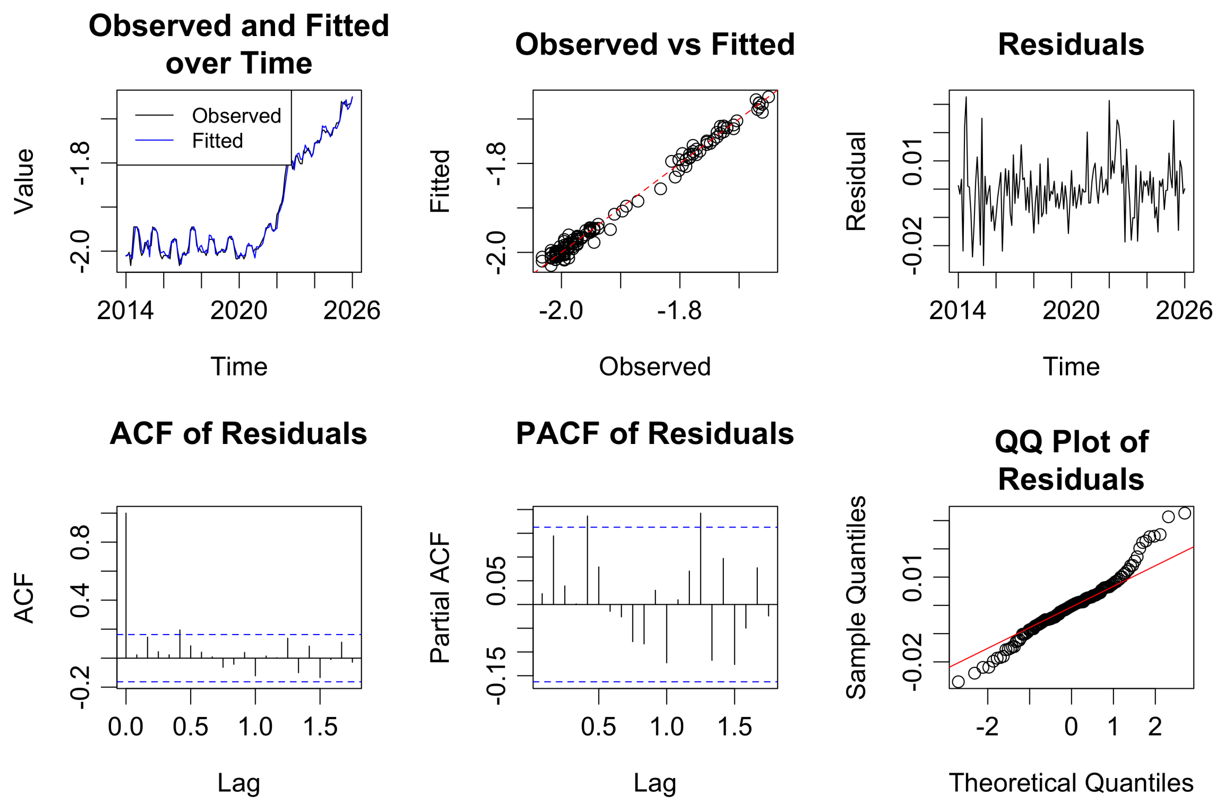


Figure 10: Model 2 Diagnostics: Despite the inclusion of external variables, residual patterns remain nearly identical to the univariate baseline.

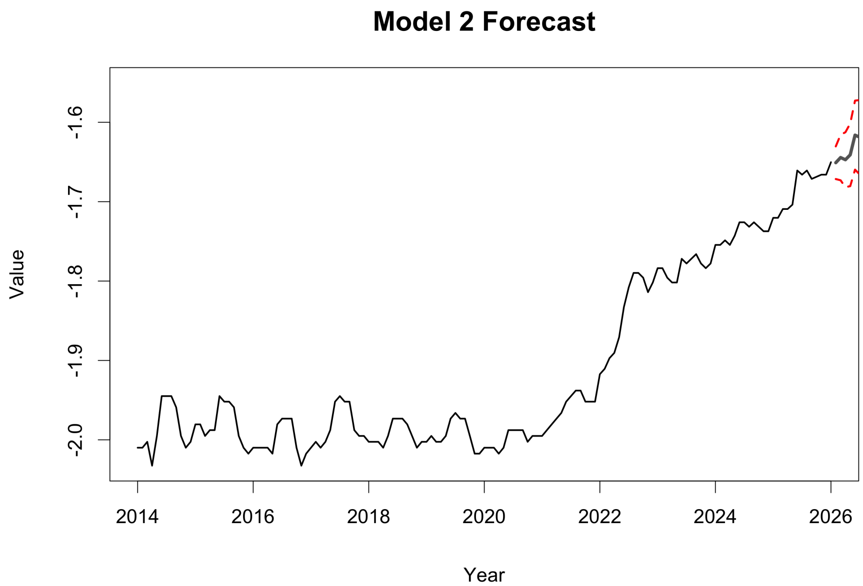


Figure 11: Model 2 forecasts electricity prices likely increasing over next 12 months.

Model 3 Diagnostics

Model 3 extended the baseline by incorporating one-month-lagged exogenous variables: a centered time trend, log-transformed Google AI Trends, and log-transformed data center investment. Upon including these predictors, the model structure transitioned to a regression with $\text{ARIMA}(1, 0, 2)(2, 0, 0)_{12}$ errors.

Table 3 shows the estimated model coefficients and standard errors. Estimates are largely consistent with those for model 2, particularly with respect to which coefficients are significant. Of course, model 3 also includes moving average terms at lags 1 and 2, but they are relatively small in magnitude and not crucial for model performance.

Seen in Figure 12, diagnostic plots are consistent with Model 2, as the observed vs. fitted correlation is 0.997, residuals are approximately white noise, and the QQ plot shows near-normal behavior with some heavy tails. Training RMSE is 0.0096 and MAPE is 0.388%. Because Models 1, 2, and 3 differ in their differencing structure and sample size due to lagging, AIC values are not directly comparable across the three models. Within the sample window common to Models 2 and 3, the error metrics are nearly identical, and neither model offers a meaningful improvement over the baseline in terms of fit.

Seen in Figure 13, the 12-month forecast is similarly reasonable, though the uncertainty intervals are also wider than the baseline given the added complexity of forecasting the external regressors forward. This is a meaningful practical limitation, as even setting aside the statistical insignificance of the external predictors, Model 3 produces less precise forecasts than the baseline precisely because it depends on variables that are themselves unknown at forecast time. The three forecast plots reinforce the case for the baseline model; it is not only the most parsimonious, but also produces the tightest and most reliable prediction intervals.

Table 3: Model 3 Coefficients: AI interest and data center investment remain statistically insignificant.

Variable	**Estimate** ($\hat{\beta}$)	**S.E.**	**t-statistic**
Autoregressive (1)	0.9498	0.0241	39.41
Moving Average (1)	0.1820	0.0931	1.95
Moving Average (2)	0.2620	0.1055	2.48
Seasonal AR(1)	0.4950	0.0886	5.59
Seasonal AR(2)	0.3228	0.0940	3.43
Intercept	-1.8756	0.0834	-22.49
Linear Time Trend	0.0369	0.0119	3.10
Log AI Trends (Lag 1)	0.0014	0.0042	0.33
Log Data Centers (Lag 3)	-0.0087	0.0080	-1.08

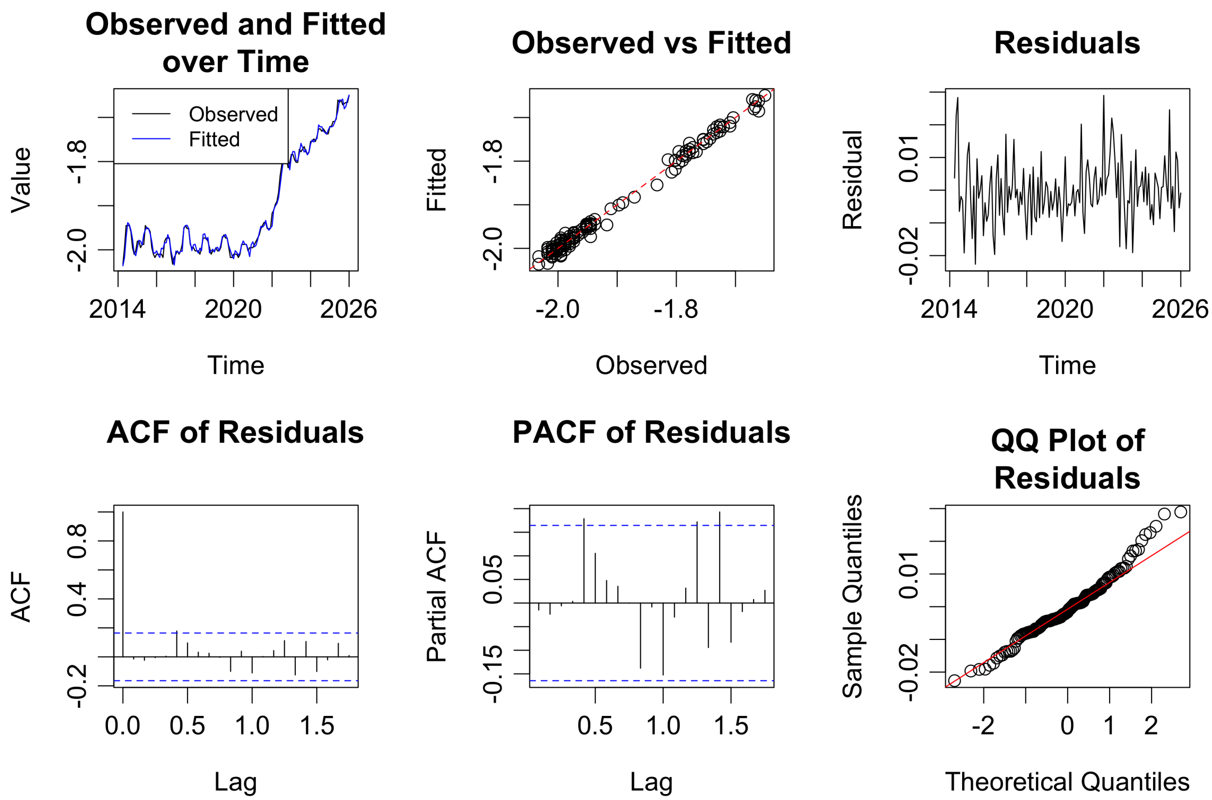


Figure 12: Model 3 Diagnostics: The residual patterns for model 3 are nearly identical to those of model 2.

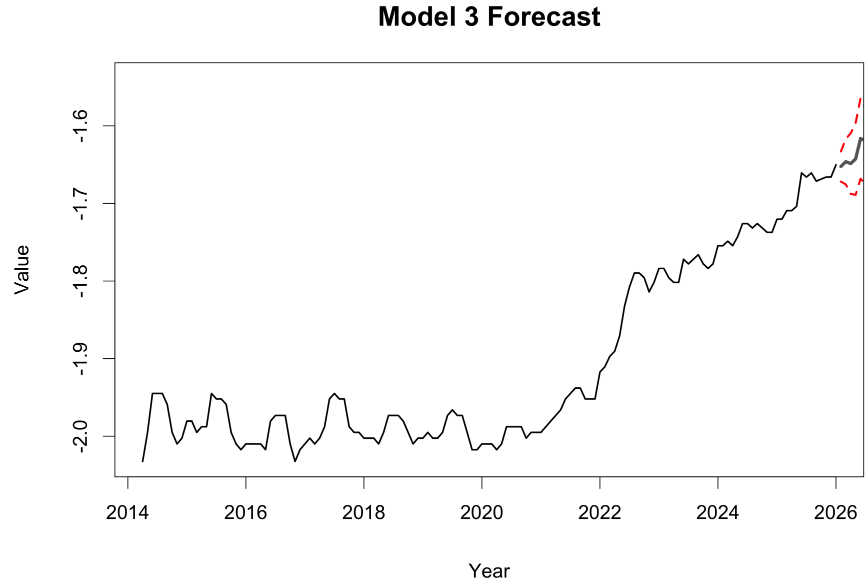


Figure 13: Model 3 forecasts electricity prices likely increasing over next 12 months.

Final Model Selection

Model	**RMSE**	**MAE**	**MAPE (%)**
Baseline SARIMA$(2,1,0)(0,1,1)_{12}$	0.00942	0.00708	0.373
Model 2: ARIMA$(1,0,0)(2,0,0)_{12}$	0.01024	0.00750	0.394
Model 3: ARIMA$(1,0,2)(2,0,0)_{12}$	0.00958	0.00737	0.388

Table 4: In-sample fit metrics for all three models. Lower values indicate better fit.

Table 4 summarizes the in-sample fit metrics across all three models. The baseline SARIMA$(2,1,0)(0,1,1)_{12}$ achieves the lowest RMSE and MAPE of the three, despite being the most parsimonious. Models 2 and 3 both include external regressors intended to capture the influence of AI-related activity and data center investment on electricity prices, but neither produces a meaningful improvement in fit. The MAPE values for Models 2 and 3 are slightly higher than the baseline, and the marginal reduction in RMSE for Model 3 relative to Model 2 does not justify the additional complexity introduced by the lagged regressors. None of the external regressors, log AI trends, log data centers, or their lagged counterparts, are statistically significant in either Model 2 or Model 3, with the time trend being the only significant exogenous predictor in both cases. This suggests that the variation in AI search interest and data center investment does not contribute explanatory power beyond the trend and autoregressive structure already captured by the baseline model. It should be noted that because the three models differ in their differencing structure and, in the case of Model 3, effective sample size due to lagging, AIC values are not directly comparable across models. Model comparison is therefore based on RMSE, MAE, and MAPE computed over each model's estimation window, along with residual diagnostics and forecast plausibility. On all of these criteria, the baseline model is preferred: it achieves the best or near-best error metrics, produces clean residuals, and generates forecasts without requiring assumptions about future values of external predictors.

Discussion

In this project, we explored the average electricity price in the U.S. over the past 12 years, with the intent of understanding its behavior and being able to predict how it will evolve in the near future. In doing so, we also looked at two other datasets, the investment amounts in data centers and the Google trend popularity for the topic AI. We utilized these two secondary datasets to uncover patterns not solely observed in the electricity prices data, with the hopes of constructing more accurate predictions after accounting for such exogenous factors.

However, the results of our analyses were not what we had entirely hoped. With regards to incorporating data center investments and Google AI trends data for more insight into electricity prices, the results of our cross-correlation analysis revealed that electricity prices neither lead nor lagged either of the other two time series. In other words, we were unable to identify a clear relationship for electricity prices with data center investments or Google AI trends data in any way. We also found that the best-performing model out of the three proposed model frameworks was the baseline case, which was fitting a seasonal ARIMA model to the electricity prices data. This is further evidence that the data centers and trends data were not significantly related to the electricity prices data, as we would expect that including cross-correlated exogenous factors would result in a model with a better fit and lower MAPE.

Limitations and Next Steps

It is important for us to address the identified limitations of our project and propose next steps for a better exploration in the future. For one, we believe that the time range in which we conducted our analysis could have introduced some issues. Looking at the original time series plots of the data, we can see that all three exhibit a rather drastic change in the overall trend in the early 2020s. For the electricity prices data, we observe an almost flat trend with the expected seasonal oscillations from 2014 to 2021, from which we observe an increasing trend with a couple spikes. We can observe similar phenomena in the data centers in Google AI trends data where the trends defined from 2014 onward abruptly changed to a more increasing trend following the introduction of ChatGPT 3.5 in late 2022. Violating the fundamental assumption of time series forecasting which is that we expect the trend and underlying data generating process to not change, we must consider the possibility that these abrupt changes biased our parameter estimates and statistical inference. Although this exploration considered the largest possible time range for which the three time series intersected in order to maximize the sample size for parameters estimates, a next step would be to conduct the same analysis on a subset of the data (e.g., 2021 onward) and see if the model outcomes and results may be any different.

Another limitation to consider is the nature of the Google AI trends data that we worked with. In particular, the issue with the data is that they do not accurately reflect search counts, rather it measures popularity relative to the total number of searches being done on Google at the time. According to the Google News Initiative,"A downward trend on the graph means a search term's relative popularity is decreasing, not necessarily that the total number of searches is decreasing" [2]. This contradiction poses issues since we would expect that the total number of searches for the topic of AI is a good measure of public interest towards the AI, however, the data we have may not reflect that properly. Another shortcoming of the Google AI trends data is its granularity and support, with the measure ranging from 0 to 100. This is particularly an issue up to 2021 where the values largely ranged from 0 to 5, with the smallest possible increment being 1. This led to a log-differenced time series plot that looks almost artificial, exhibit the same peaks and troughs repeatedly. Lacking the level of granularity and precision that would allow for a good model fit to the data, a next step would be to discover a data source that more closely reflects actual search counts for the topic AI and is more continuously-valued.

One final next step would be to consider additional data sources that could be linked to electricity prices. Our intuition was that the exploding public interest in AI and the rapidly growing investments into data centers would put upwards pressure onto electricity prices, forcing them to increase in the last couple of years. However, our analyses showed that this might not be the case, with no significant cross-correlations observed for electricity prices with data center investments and Google AI trends data. Although we observed that electricity prices are largely driven by their past observations and trends, contextually, it is non-sensical to believe that they are completely isolated. After all, we observed that the spikes in electricity prices coincided with increasing fossil fuel prices from the Russia-Ukraine war and Trump's Liberation Day tariffs. Thus, we believe that there should exist other time series data that is cross-correlated with electricity prices, such as fossil fuel prices or inflation rates. One final next step would be to use context and our reasoning to identify these other time series data that are cross-correlated with electricity prices, with the intent of ultimately producing a very accurate and robust prediction model that does not solely rely on its past observations to make forecasts.

References

[1] Google Trends. *Artificial Intelligence Search Interest Over Time*. 2026. URL: https://trends.google.com/explore?q=%2Fm%2F0mkz&date=all&geo=US.

[2] Google Trends. *Google Trends: Understanding the data*. 2026. URL: https://newsinitiative.withgoogle.com/resources/trainings/google-trends-understanding-the-data/.

[3] Ryan Bellefontaine. *Visualized: U.S. Data Center Investment (2014–2025)*. 2026. URL: https://www.visualcapitalist.com/sp/bhp01-visualized-u-s-data-center-investment-2014-2025/.

[4] U.S. Bureau of Labor Statistics. *Average Price: Electricity per Kilowatt-Hour in U.S. City Average*. 2026. URL: https://fred.stlouisfed.org/series/APU000072610.