MSCA Time Series Analysis: PJME Energy Consumption Prediction

Richard Yang

Agenda

1 — 2 — 3 — 4

1. Business Problem

2. Data Overview & EDA

- ❑ Data source
- ❑ Data cleaning
- ❑ Seasonality
- ❑ Hypothesis Test

3. Model Approaches & Comparison

- ❑ Baseline (LR + TSLM)
- ❑ Exponential Smoothing
- ❑ Dynamic Harmonic Regression
- ❑ sARIMA
- ❑ Bayesian Structural Time Series

4. Conclusion & Recommendations

Why is Energy Consumption Critical?

Important Macroeconomic Indicator	Households Financial Pressure	Unstable Consumption Quantity
Following the pandemic, there has been a surge in energy consumption, leading to elevated prices. This, in turn, has contributed to heightened inflation rates and triggered economic recession.	About 20% of U.S. households have missed or made a late payment on their utility bill in the last month. Families with an income of $50,000 or less are struggling the most to absorb higher energy costs.	The energy sector is characterized by significant fluctuations due to supply-demand imbalances, geopolitical tensions, and other macroeconomic factors. Therefore, it is very challenging to construct a robust model to provide accurate predictions.

Source

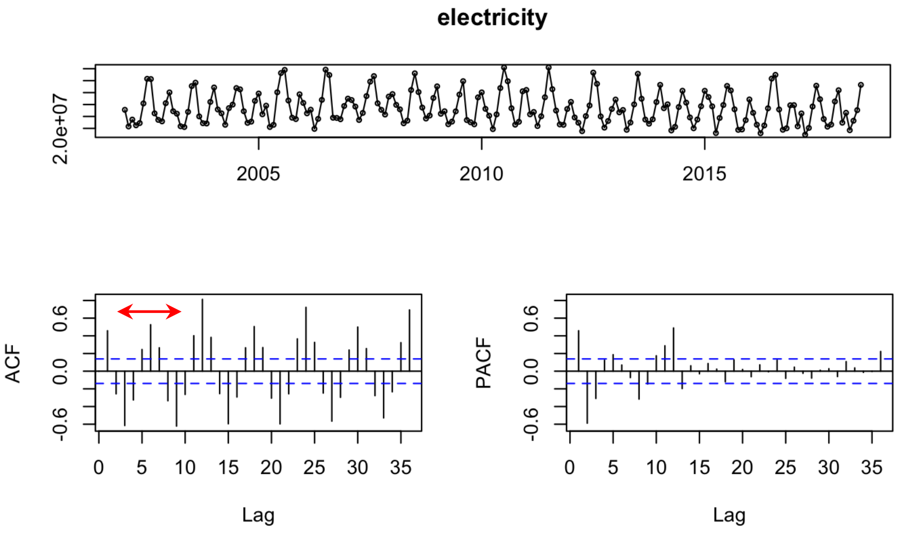
Benefits of Understanding Trend

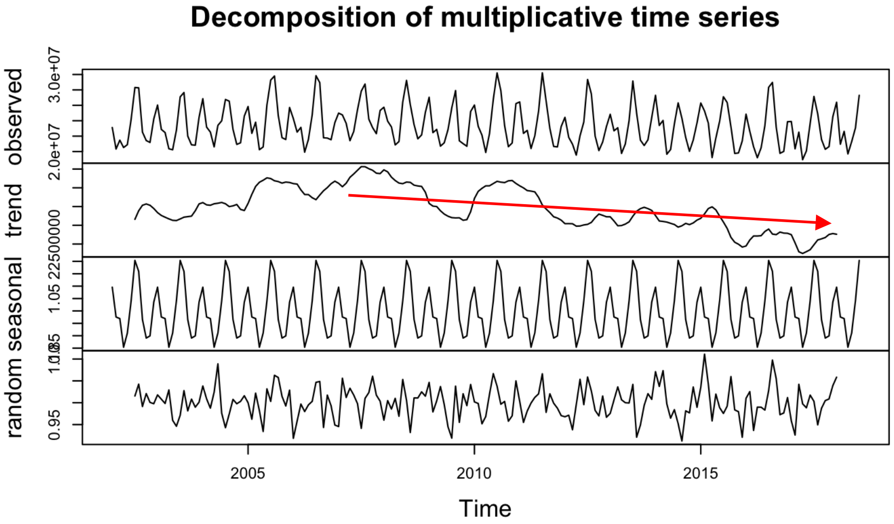
01	**Cost Savings**	• Optimize purchase decisions. • Minimize operational expenses. • Boost overall profitability.
02	**Strategic Planning**	• Align operations with price trends. • Maximize productivity during low costs. • Enhance financial forecasting accuracy.
03	**Competitive Advantage**	• Outperform market peers. • Enhance brand trustworthiness. • Secure favorable contract terms.
04	**Risk Mitigation**	• Hedge against price swings. • Reduce financial exposure. • Proactively manage market volatility.

Data Overview

1	**PJM Interconnection LLC (PJM)**	Contains hourly energy consumption from 2002 to 2018
2	**Data Size**	2 columns (energy & time), 145k hourly observations -> 200 months (after transformation)
3	**Data Cleaning & downsampling**	Inconsistent, missing data at certain hour/month (e.g. 2008-March massive missing data)
4	**Train / Test Split**	Train 75%, Test 15% (36 months)

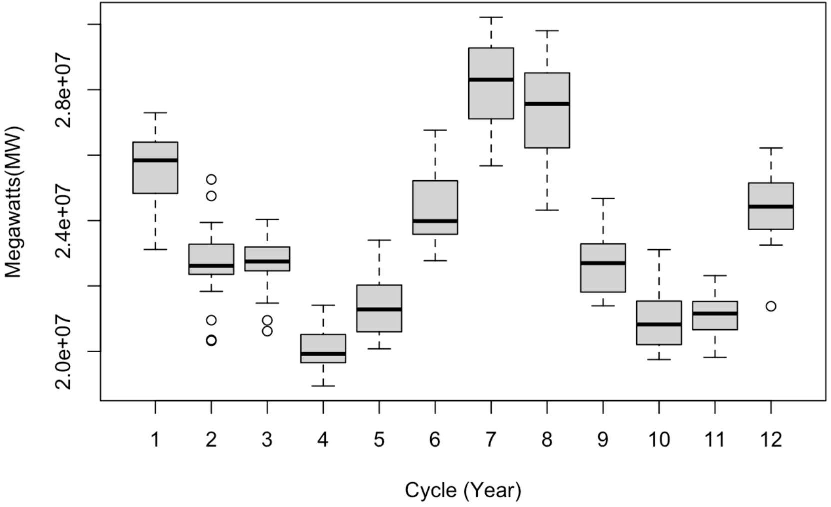
EDA: ACF, PACF, Trend

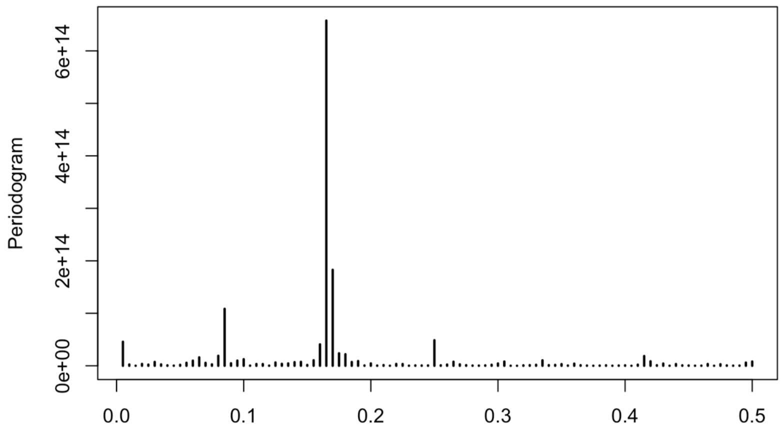


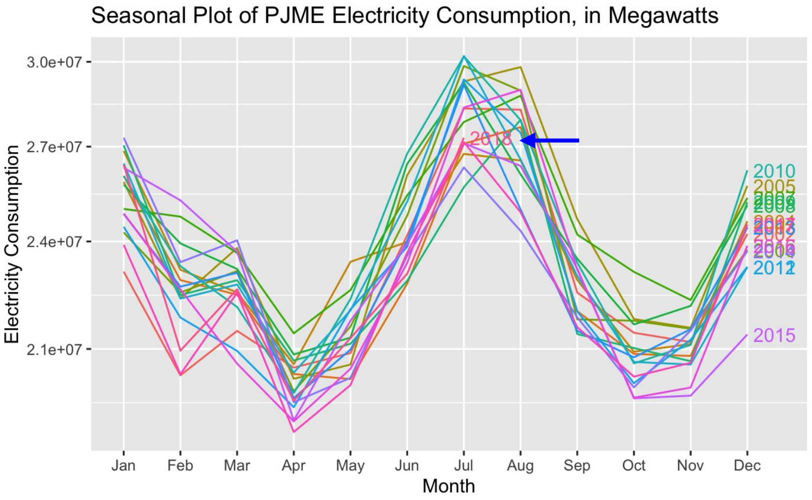


- The PJME monthly average electricity consumption has:
  - a seasonal pattern (six month)
  - a declining trend (also see U.S. Energy Information Administration report)

EDA: Decomposition, Spectral Analysis







- **Periodogram** - Dominant frequencies t = 0.165
- **Boxplot** - Semi-annual electricity consumption peaksummer (June / July) and winter (Dec / Jan)
- **Seasonal plot** - echoes boxplot, but also shows variation

EDA: Hypothesis Test

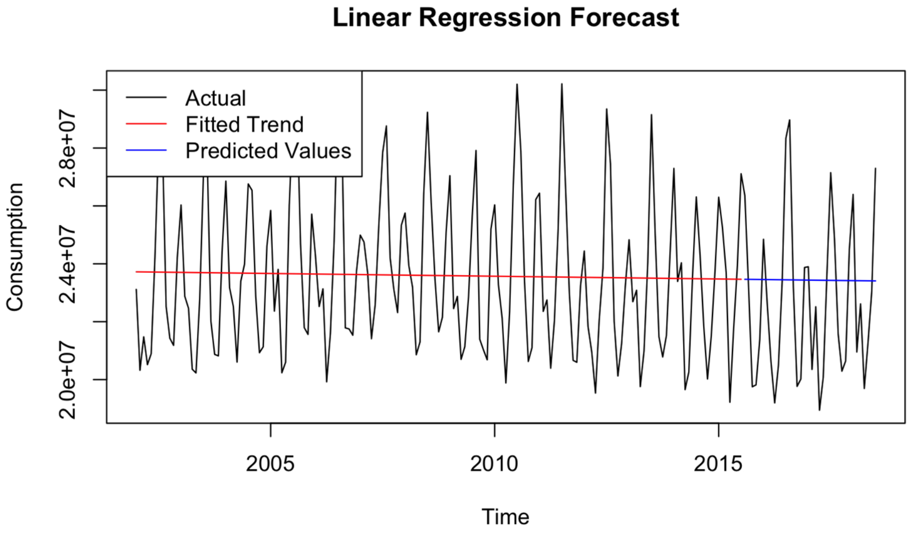
01	Stationarity	KPSS Level = 0.44894, Truncation lag parameter = 4, p-value = 0.05606
02	Differenced Stationarity	KPSS Level = 0.024273, Truncation lag parameter = 4, p-value = 0.1
03	Stability	Box-Cox Lambda = 0.2452849

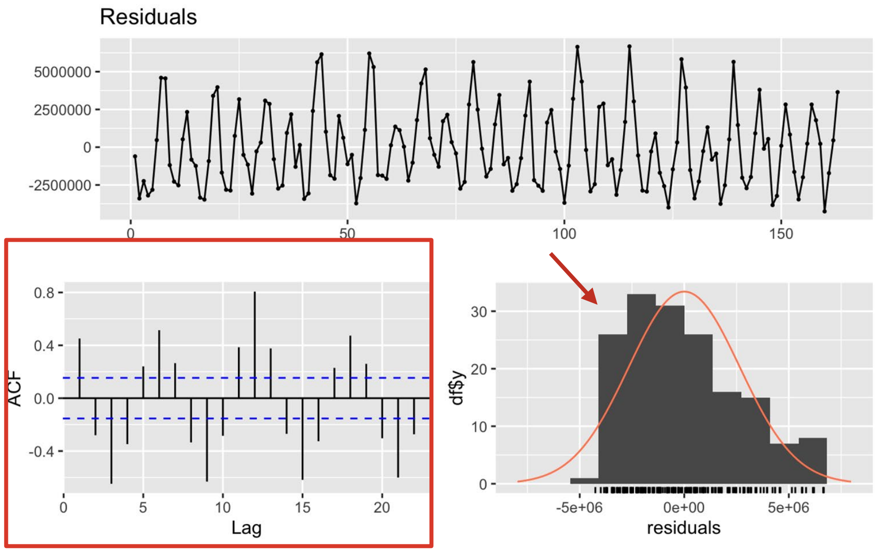
- **KPSS Level Stationary Test** - $p > 0.05$, we cannot reject the null hypothesis which means that our data can be concluded as stationary
- **Box-Cox Transformation** - a very slight downward shift with the level of the series

Linear Regression (LR): Split Validation

RMSE	2821861
MAE	2391181

Since LR ignored seasonality, the model clearly failed to capture seasonal variations and we have residuals with seasonality as well

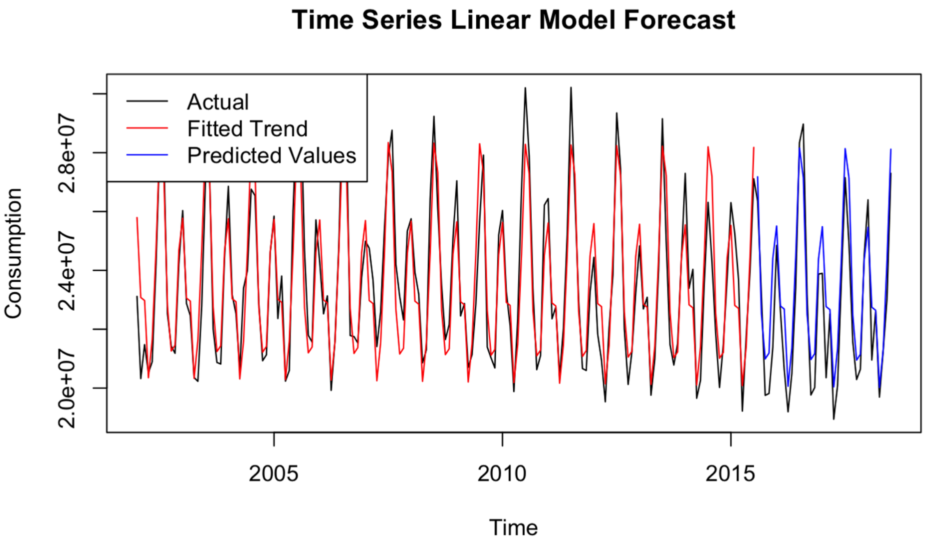


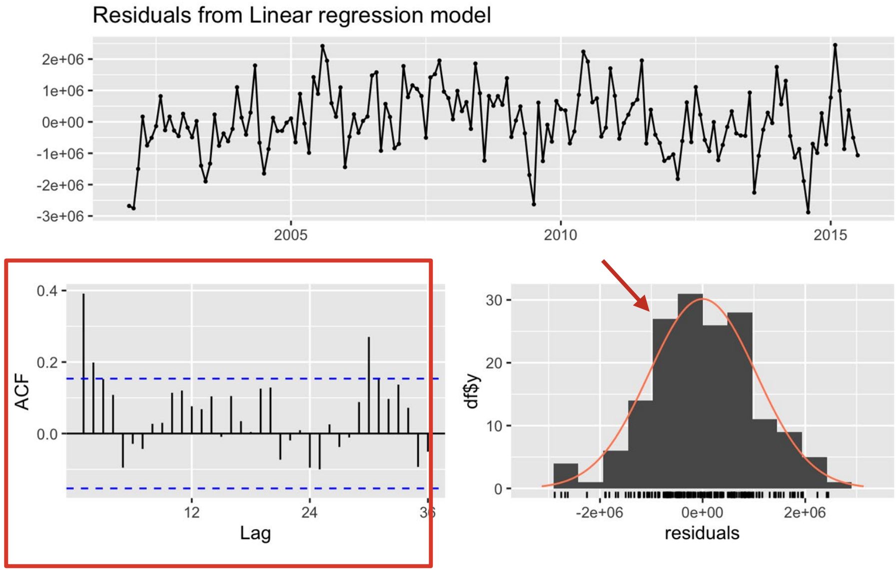


Time Series Linear Model (TSLM): Split Validation

RMSE	1220888
MAE	1007208

TSLM is also experimented to incorporate seasonality. The prediction significantly improves and the residuals mostly resemble a white noise, which is a good improvement as a baseline model





ETS: Exponential Smoothing

- Additive Holt -Winters

Additive Holt-Winters Formulas:

1. **Level (l) Update:**

$$l_t = \alpha * (y_t - s_{t-m}) + (1 - \alpha) * (l_{t-1} + b_{t-1})$$

2. **Trend (b) Update:**

$$b_t = \beta * (l_t - l_{t-1}) + (1 - \beta) * b_{t-1}$$

3. **Seasonality (s) Update:**

$$s_t = \gamma * (y_t - l_t) + (1 - \gamma) * s_{t-m}$$

4. **Forecast (F) at Time t+k:**

$$F_{t+k} = l_t + k * b_t + s_{t+k-m}$$

Using additive Holt-Winter since the variance is relatively stable

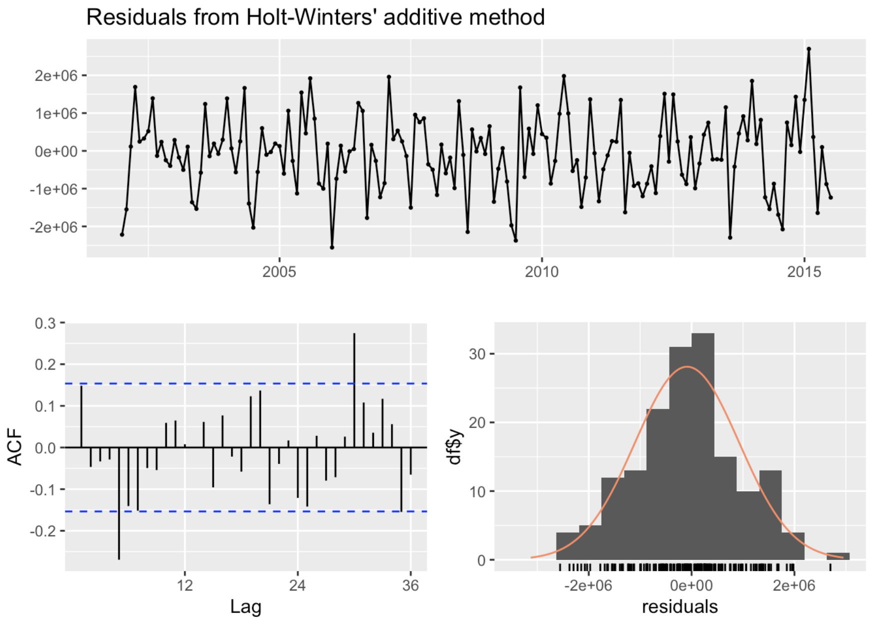
Smoothing parameters

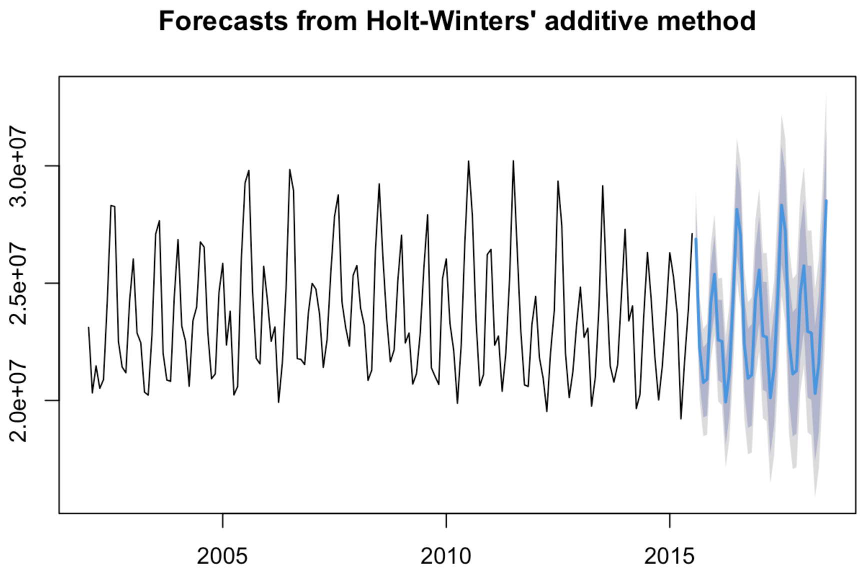
alpha	0.3042
beta	0.0015
gamma	1e-04

Initial states

l	23163572.545
b	36414.1843
s (m=12)	926106.6, -2307248, -2434922, -974162.2, 3708332, 4814147, 1015599, -2084154, -3355885, -755715.7, -685822.6, 2133724

ETS: Evaluation of HW Model





The ACF plot seems to have autocorrelation issue

AIC	5360.97
AICc	5365.191
BIC	5413.564

ETS: Fix Autocorrelation - ARIMA with CV

RMSE of CV[5]

	X1	X2	X3	X4	X5
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
pdq_000	1005448.5	995820.6	943605.0	1141727	820116.2
pdq_001	1007967.6	994032.2	941356.4	1142022	850624.2
pdq_005	996817.9	984310.8	950269.1	1125259	420180.4
pdq_010	1150823.3	1224274.6	1576908.6	1177413	1123050.1
pdq_011	1272167.0	995820.6	943605.0	1141727	820116.2
pdq_015	1000094.4	994797.3	943771.9	1135941	1002262.8
pdq_100	1014304.7	994465.6	941704.3	1142531	831848.7
pdq_101	1007873.3	995035.1	942045.1	1141448	854995.4
pdq_105	993365.3	995225.5	931363.6	1141581	505900.2
pdq_110	1265511.4	1193353.1	1305312.6	1309862	598966.4
pdq_111	1016403.3	994343.0	941457.4	1142603	832202.1
pdq_115	1019667.7	993704.8	954402.8	1145174	805195.3
pdq_500	1054066.0	994275.2	962395.1	1121991	503731.5
pdq_501	1119659.9	1030630.2	959766.1	1156486	540872.8
pdq_505	980150.4	1001204.8	930326.2	1151843	430033.4
pdq_510	1249422.5	1211551.0	1024457.8	1245238	1146245.1
pdq_511	1058550.0	994684.1	962159.9	1123038	506101.6
pdq_515	1026594.2	1001341.9	990305.6	1152183	383336.7

- **Apply a five-fold expanding window approach to cross-validation.**
- **Aggregate the windows of RMSE to decide best ARIMA model**

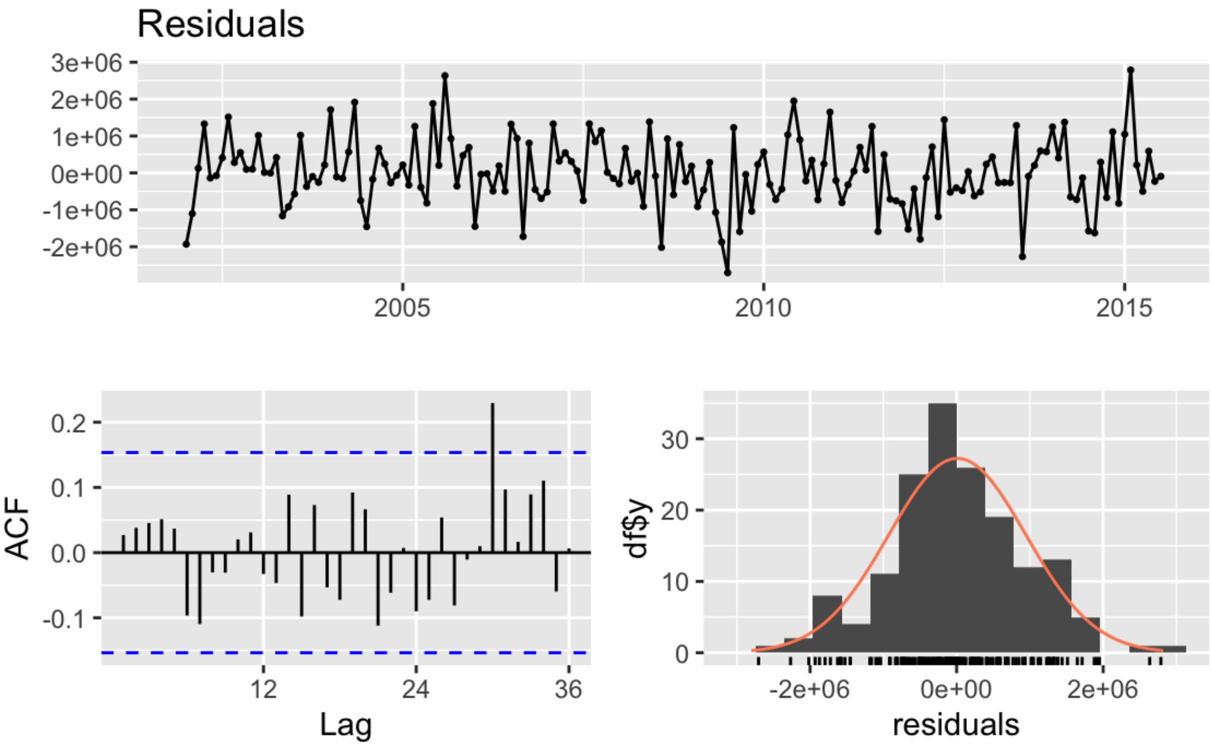
ARIMA(0,0,5)

```
Series: residuals
ARIMA(0,0,5) with non-zero mean

Coefficients:
         ma1      ma2      ma3      ma4      ma5       mean
      0.0648  -0.2043  -0.1829  -0.1044  -0.3553  -87666.94
s.e.  0.0775   0.0707   0.0791   0.0820   0.0753   17967.43

sigma^2 = 9.048e+11:  log likelihood = -2472.65
AIC=4959.31   AICc=4960.03   BIC=4980.96
```

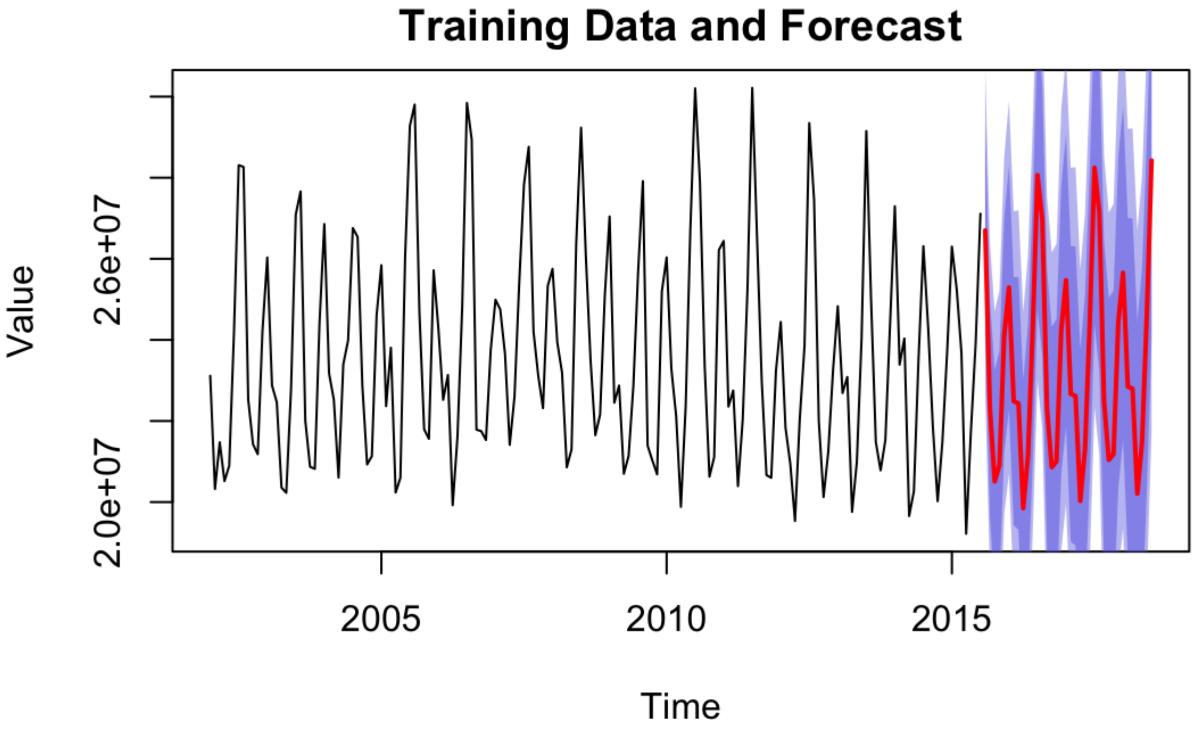
ETS: Residual after HW + ARIMA(0,0,5)



- **Less autocorrelated**
- **Residuals ~ Normal distribution**

Ljung-Box test
p-value: 0.6463

ETS: Prediction of Test Set



Test Set RMSE/MAE

HW method only

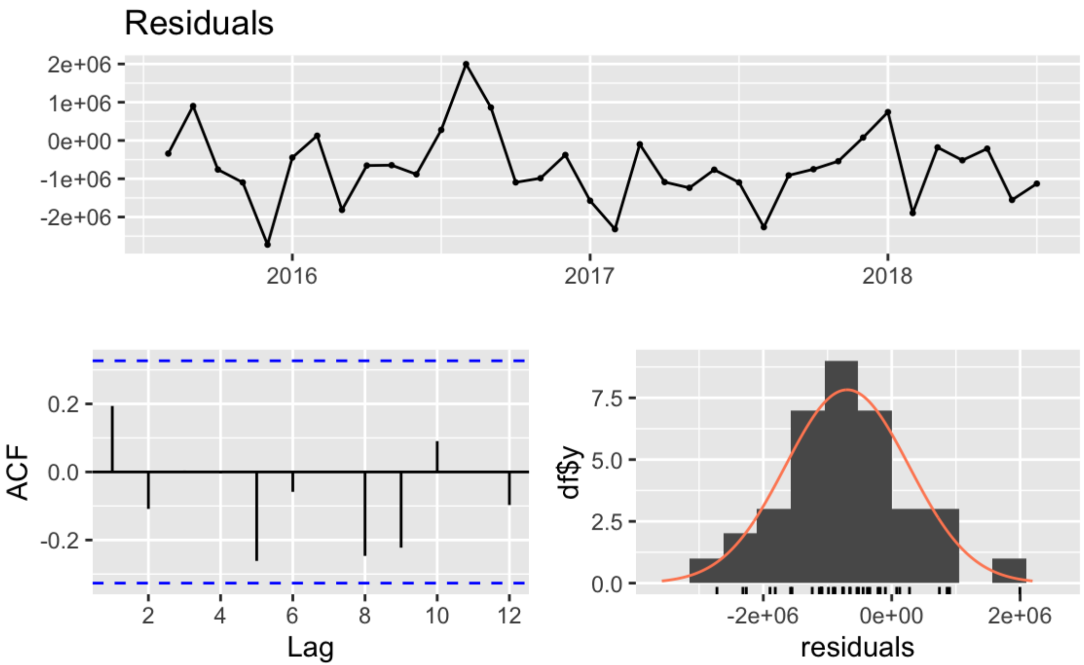
RMSE	1232816
MAE	1033276

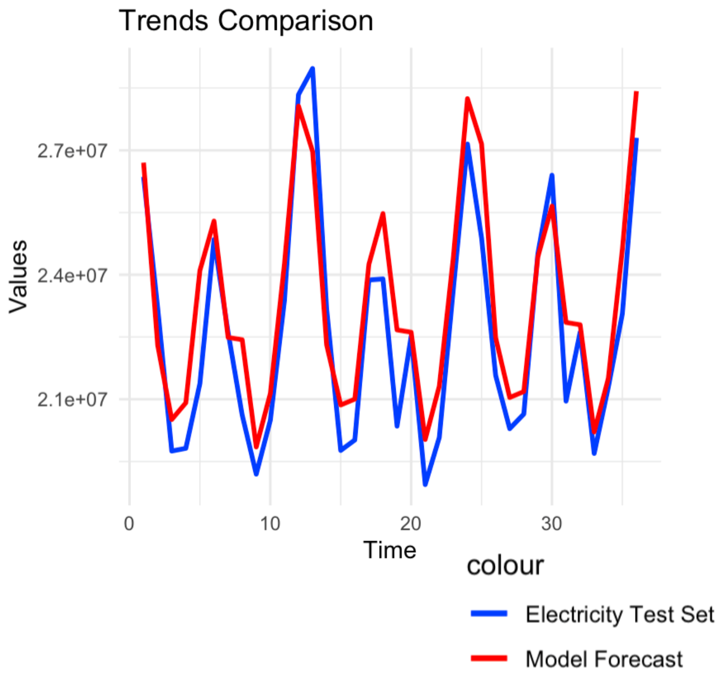
HW method + ARIMA(0,0,5)

RMSE	1175285.33
MAE	970414.86

- The confidence interval is larger than only with HW method
- The RMSE and MAE of the combined model are lower than only with HW method

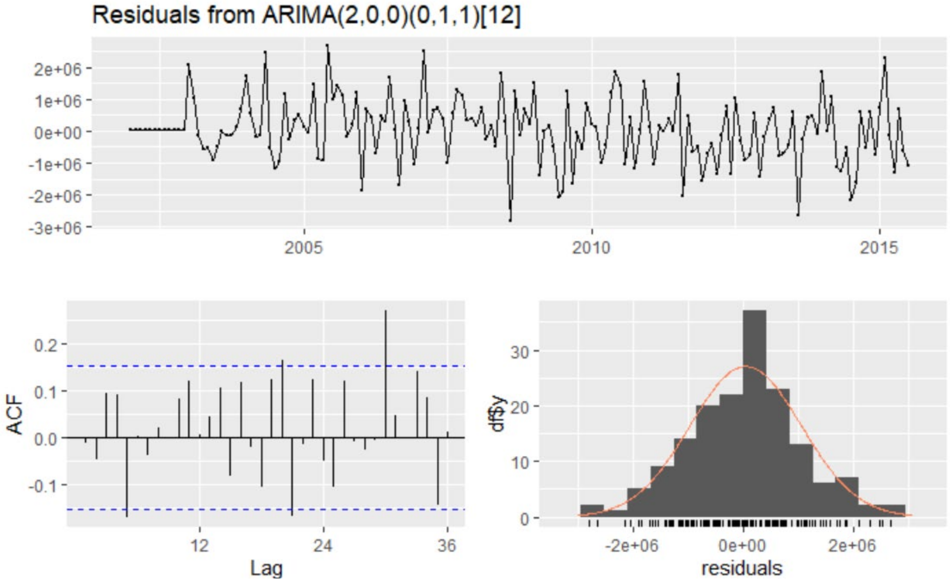
ETS: Evaluation of Test Set





- No autocorrelation issue
- Mean is not 0, which means the model might be biased
- The predicted trend follow the actual data pretty well

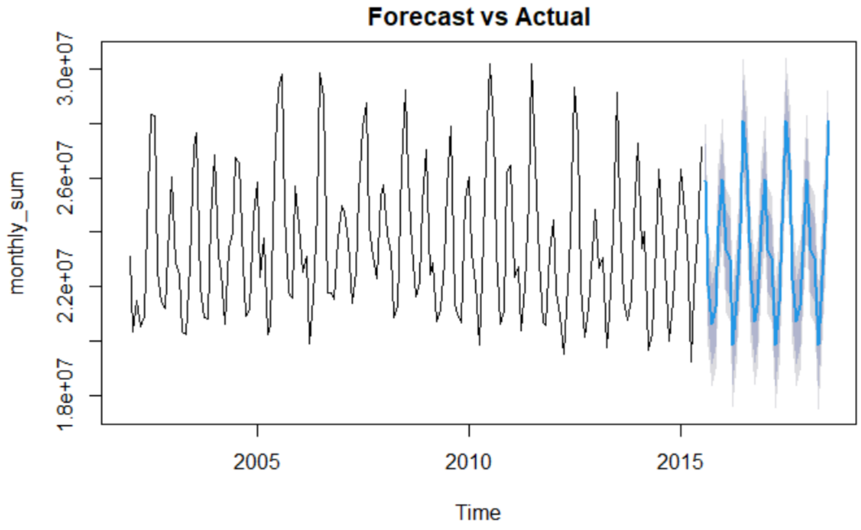
Seasonal ARIMA Model

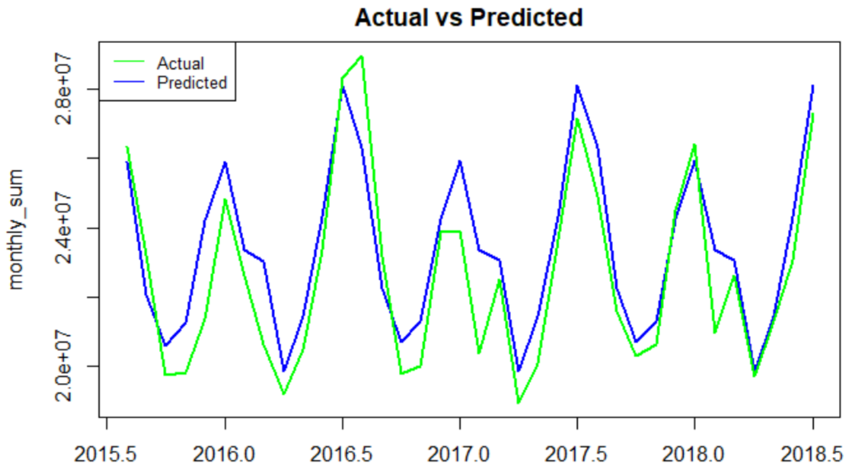


RMSE	1309595.63
MAE	1068540.59

- The ARIMA model includes an autoregressive term of **order 2**, indicating the current value is influenced by its two previous values.
- There's **no non-seasonal** differencing or moving average component.
- The model requires **first differencing** for the seasonally adjusted series **every 12 periods**.

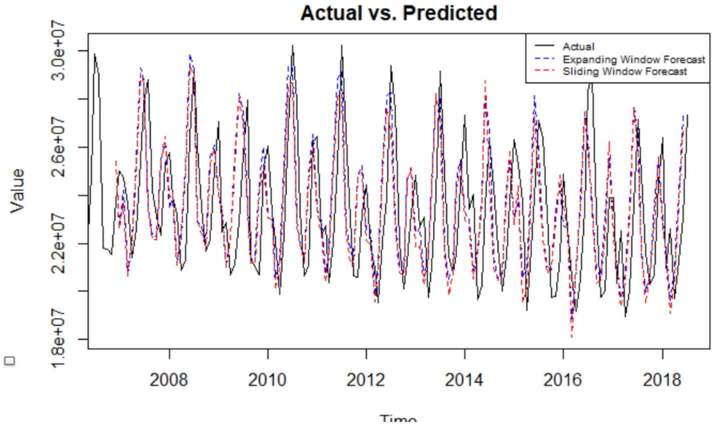
Seasonal ARIMA: Forecast Result



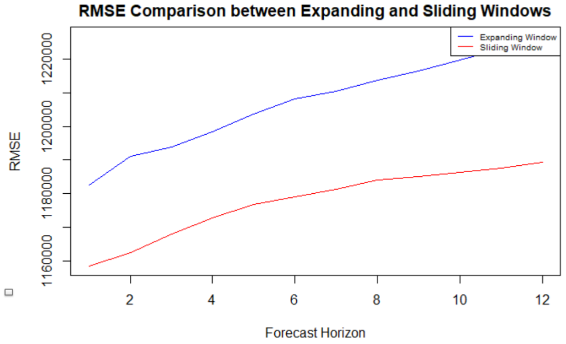


- This ARIMA Model is capable of capturing the seasonality well especially the peaks and troughs.
- However, sometimes it may become too insensitive in capturing the detailed changes.

ARIMA: Sliding & Expanding Window

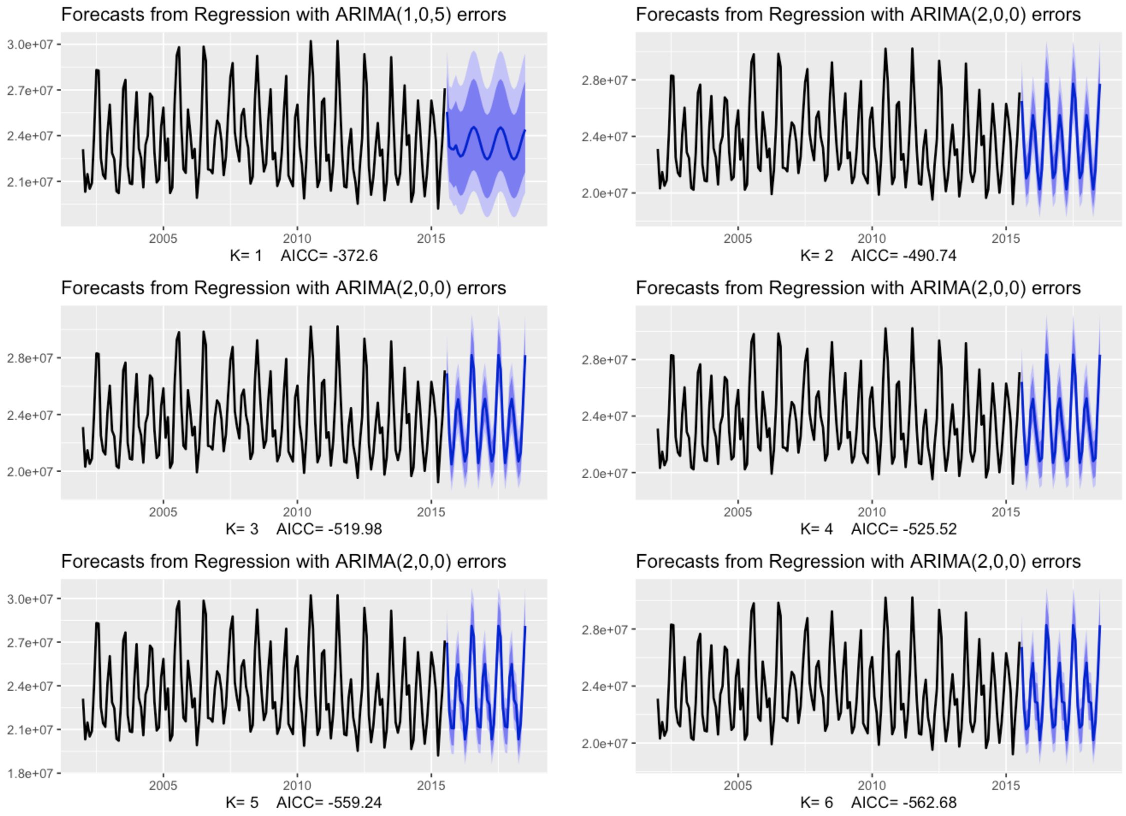


	Expanding Window	Sliding Window
RMSE	1207319	1177700
MAE	983325.3	938397.1



- The RMSE and MAE are both lower in Sliding Window ARIMA model.
- The dynamic RMSE of Sliding Model over time tends to have a more stationary trend. It is more stable.
- Older data might contain noise or anomalies that aren't representative of current trends. By focusing on recent data, the sliding window might avoid being influenced by this noise.
- Since the data has strong seasonality, the sliding window might be more adaptive. It will be more sensitive to recent patterns than an expanding window.

DHR: Dynamic Harmonic Regression with ARIMA errors

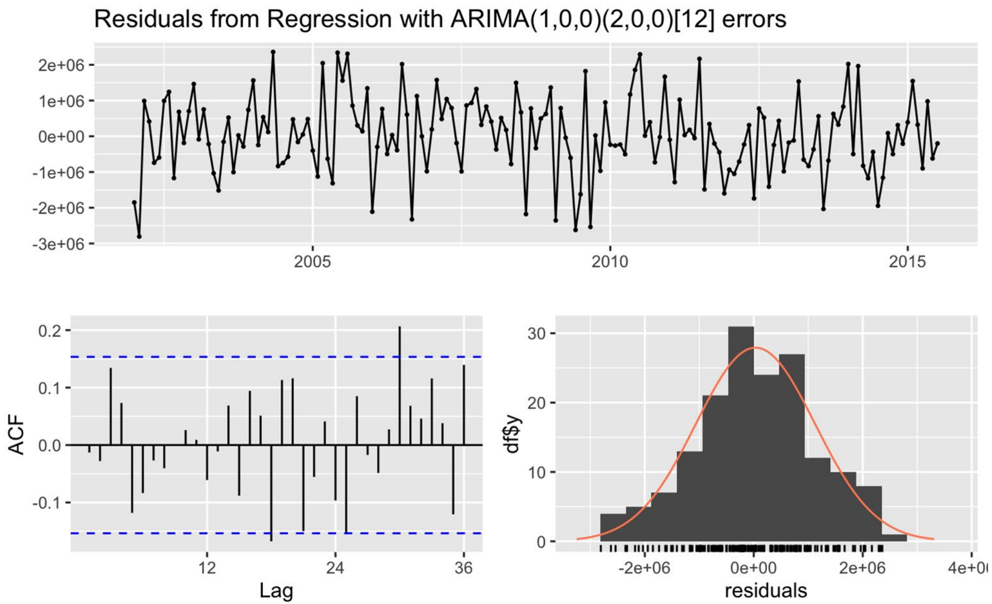


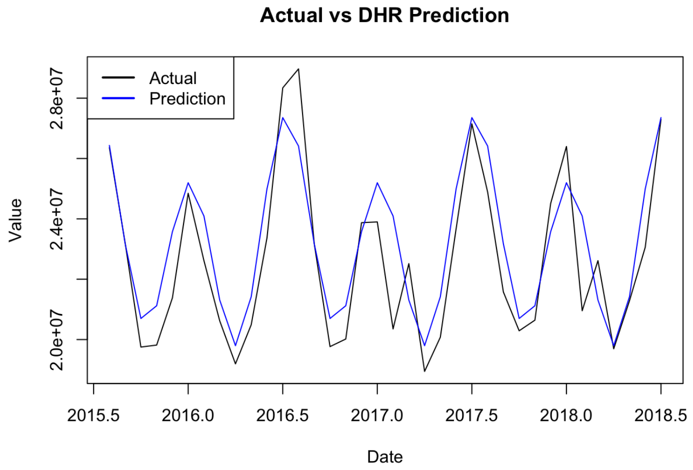
Best K	2
Best ARIMA param	(2,0,0)
Auto ARIMA	(1,0,0)(2,0,0)[12]

DHR: Split Validation

RMSE	1540094.27
MAE	1177421.18

- The prediction result is much worse but the residuals follow the pattern of a white noise with most ACF values staying within the boundaries
- Daily vs monthly: trade off of data, noise, and model memory





BSTS: Bayesian Structural Time Series

Mathematics behind BSTS

Basic equation: $y_t = \mu_t + \tau_t + \alpha_t + \epsilon_t$

Local level linear model: $\mu_{t+1} = \mu_t + \delta_t + \eta_{0t}$

Slope of Local: $\delta_{t+1} = \delta_t + \eta_{1t}$

Seasonal: $\tau_{t+1} = -\sum_{s=1}^{S-1} \tau_t + \eta_{2t}$

Autoregressive: $\alpha_t = c + \phi_1 \alpha_{t-1} + \phi_2 \alpha_{t-2} + \ldots + \phi_p \alpha_{t-p} + \eta_{3t}$

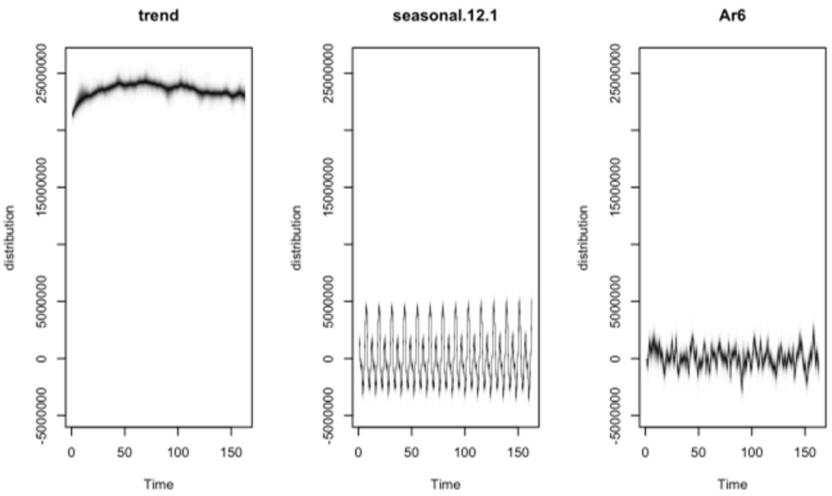
Errors: $\epsilon_t \sim N(0, \sigma^2)$ $\eta_{0t}, \eta_{1t}, \eta_{2t}, \eta_{3t} \sim N(0, \gamma^2)$

Components of this BSTS model:

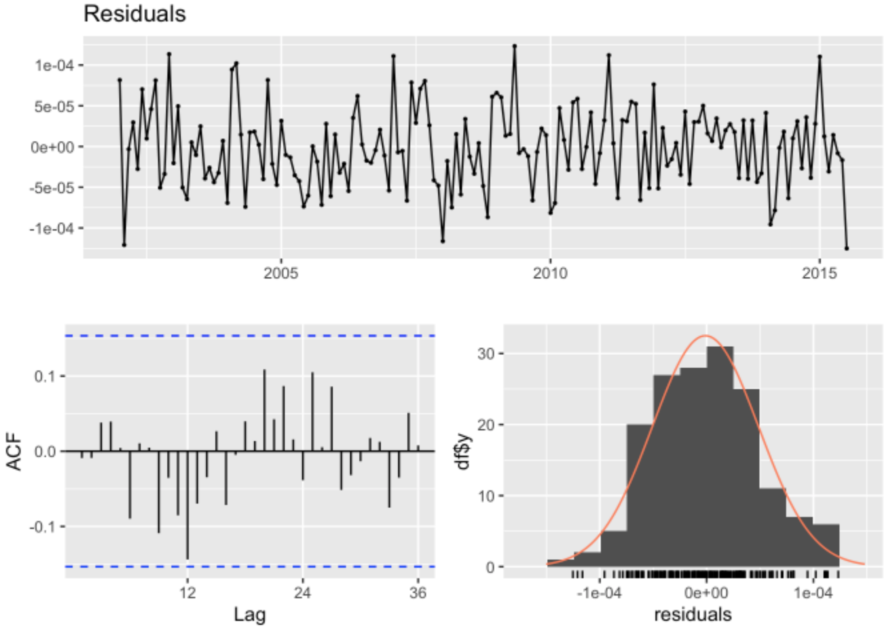
- Local Linear Trend
- Seasonal (12)
- Auto AR Process (lag=6)

Source

BSTS: Bayesian Structural Time Series



Ljung-Box test
p-value: 0.8875



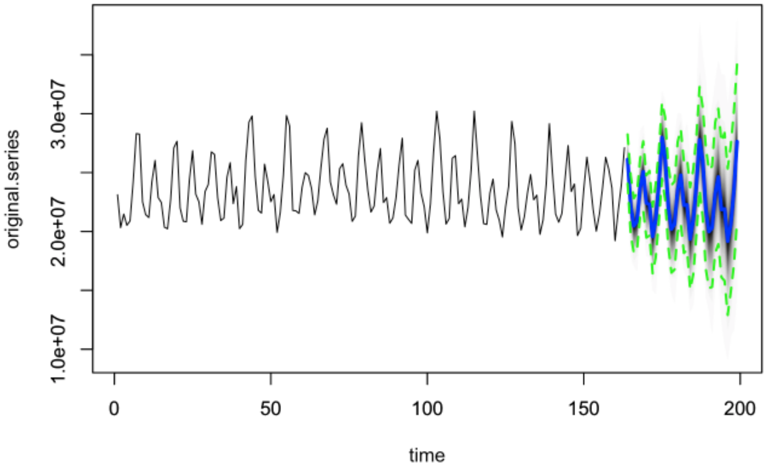
Components of this BSTS model:

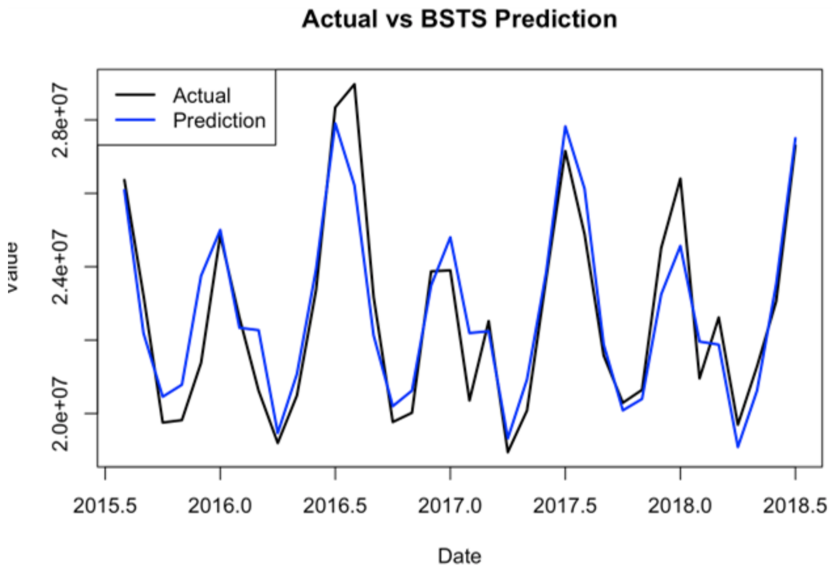
- Local Linear Trend
- Seasonal (12)
- Auto AR Process (lag=6)

BSTS: Prediction Results

The bsts forecast shows a **stochastic trend** in longer forecast horizons, allowing greater uncertainty in future growth.

RMSE	1002598
MAE	787623





Model Comparison

Why BSTS has the best performance?

- Flexibility and transparency to add components and fine tune
- Can handle non-Gaussian distribution assumptions for the errors, better at capturing outliers
- The MCMC draws from the posterior predictive distribution automatically account for model uncertainty

Why do other methods have worse performance?

- Problem of ETS method is the model for test set is biased
- DHR is more useful when addressing longer seasonal period, while our data has a shorter period of seasonality--and models like ARIMA is sufficient to handle

Models	RMSE	MAE
LR	2,821,861	2,391,181
TSLM	1,220,888	1,007,208
ARIMA	1,177,700	938,397
ETS	1,175,285	970,414
DHR	1,540,094	1,177,421
BSTS	**1,002,598**	**787,623**

Conclusion

1. Develop a predictive model for average of monthly energy consumption that accurately forecasts the energy usage pattern.
2. Enable energy providers, businesses, and households to plan for resource allocation, infrastructure upgrades, cost management.
3. Choose the BSTS model for optimal performance, minimal assumptions, and robust incorporation of uncertainty, enhancing its inferential capability

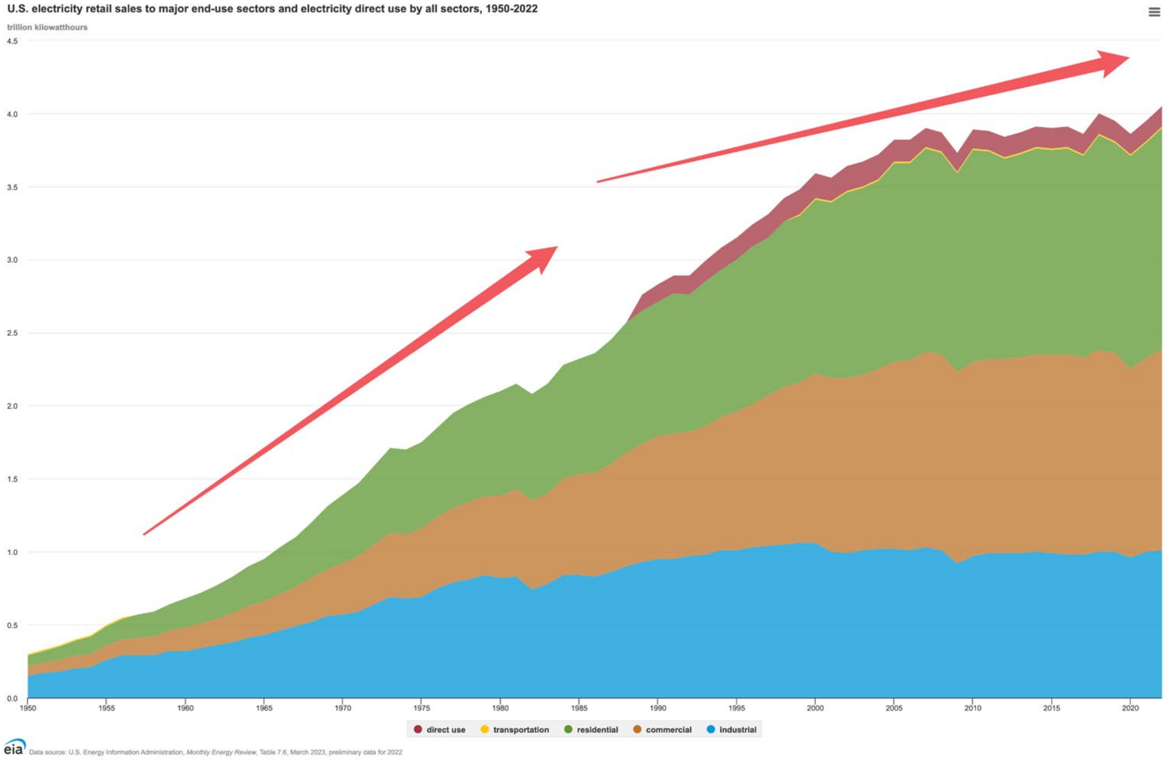
Future Works

1. Future work includes predicting consumption in different scales such as daily, hourly data.
2. Incorporate additional explanatory variables, such as specific events like holidays or weather conditions.
3. Experiment with more components with the BSTS to get better performance

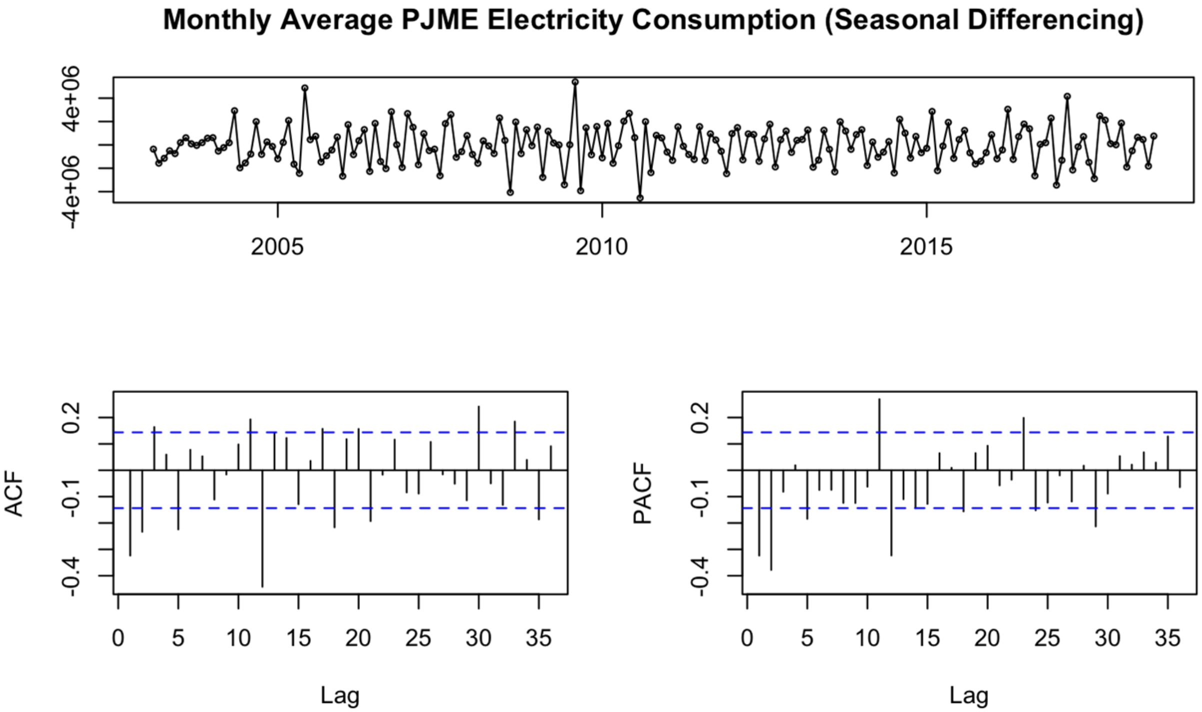
Appendix: Contribution

	ARIMA @Richard Yang	Exponential smoothing @Chenfeng Tsai	Bayesian Structural Time Series @Sam Zihao ding	Dynamic Harmonic Regression @Han Jiang
EDA (general)	Han Jiang ● Seasonality ● Trend Decomposition & Detrend ● ACF / PACF ● KPSS & Box-Cox ● Paradigm			
Model Building & Training	ARIMA & SARIMA	Holt-winters + arima for residuals	BSTS	LR and TSLM DHR with auto.arima
Slide Deck	● Modeling ● Business Problem	● Modeling ● Conclusion	● Modeling ● Model Comparison	● Modeling ● EDA & Data Overview

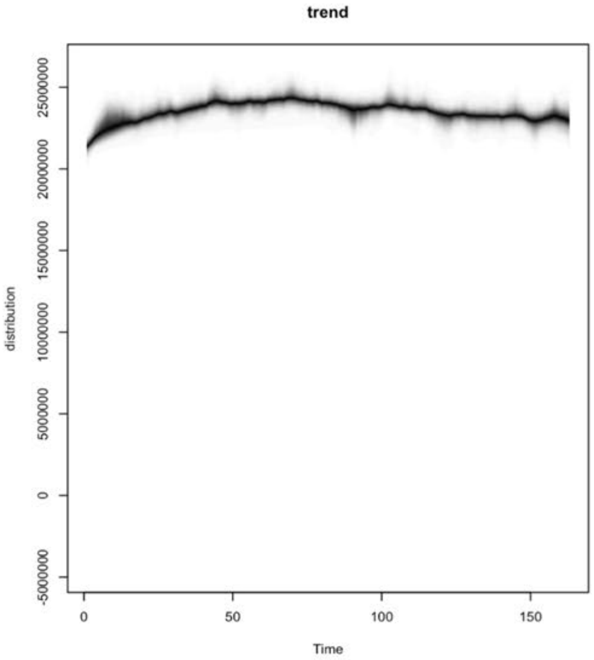
Appendix: U.S. Electricity Consumption by Sector

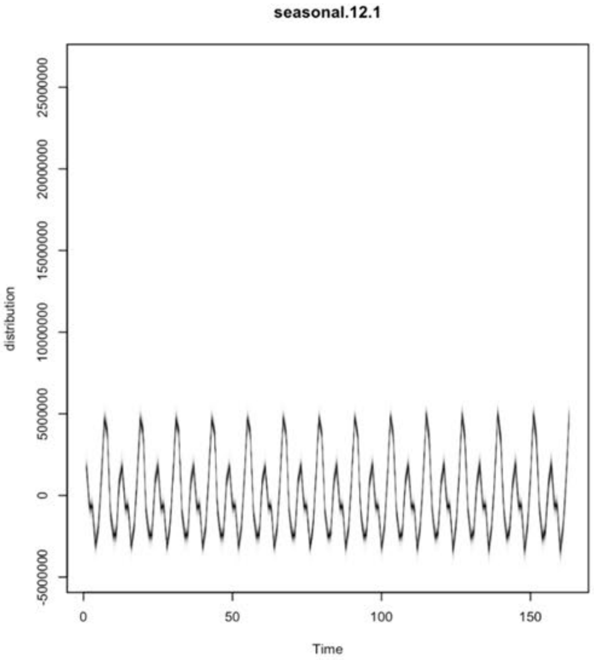


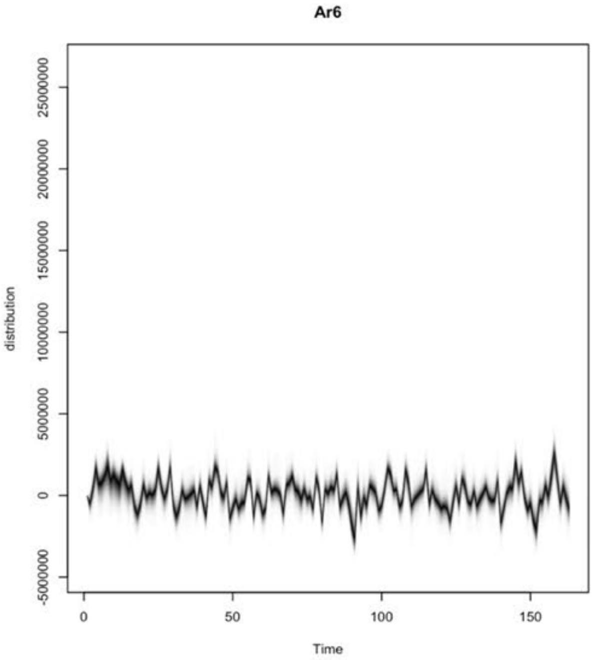
Appendix: Seasonal Differencing



Appendix: BSTS Components







Appendix: QQ Plot for BSTS