1 Review: Project 12

1.1 Strength

There are some conceptual and technical merits of authors' report about household electricity consumption. Conceptually, even though authors' report doesn't present conclusion on this topic, the topic itself is worth our concern and more study. Technically, many methods of class are applied, and the analysis is partially correct.

1.2 Weakness

In "Data Sources" section, the dataset is not representative. As shown in the dataset's description, it only contains the electricity consumption data of a single household, and, therefore, it probably can't represent the electricity consumption of any region. All we can get from the time series analysis is just the electricity consumption pattern of a single family. In this case, authors can't state that their research can "serve as crucial reference for energy policy formulation and resource allocation, fostering societal sustainable development".

Data are inconsistent in different parts of the report. In sections "Data Sources" and "EDA and Model Specification", authors analyze the daily electricity consumption but, in the following sections, they try to select model based on monthly data. When authors finally decide to select model using monthly data, their previous efforts on analyzing daily data may be useless.

In "Missing values" section, authors directly drop the dates with missing data and this may result in inconsecutive time series. Authors may need to find a better strategy for treating the missing data. The gaps in the time series data may negatively influence the ARMA model performance.

In "Detect the outliers" section, authors directly drop the outliers without enough justification. Authors should carefully consider why there are outliers. Do the outliers come from errors in recording or they are just a part of natural process? In addition, it's strange that the minimum is still in the dataset after they drop the outliers. This may be due to the program bugs.

In "Subsample the data" section, they subsample data by "selecting every other day", which enlarges the gaps in time series. We question whether authors' data is still suitable for time series analysis, because their practice of dropping missing values, outliers and every other day may seriously undermines the continuity of time series.

In "EDA and Model Specification" section, the analysis of ACF and PACF plot is wrong. For ACF plot, the phenomenon that "time series is slowly decaying" can't lead to the conclusion "the presence of significant autocorrelations between the data points". The significant autocorrelation is identified by the ACF values outside the confidence interval. For PACF plot, it's hard to see why authors argue that there are "significant lag components beyond this point". When we look at the PACF plot, most points after lag = 2 fall into the confidence interval.

In "Trend" section, they fail to find the trend using linear regression but this doesn't mean there is no trend within data. When we look at the decomposition plot in section 6.1, we can find that the trend found by

LOESS is significant. This may contradict authors' argument.

In "Spectrum Analysis" section, they draw the smoothed periodogram in a wrong way and analyze the period through eyeballing the plot. When we inspect their code, we can find that the span of smoothed periodogram is set to $(4, 6, 4)$ which is relatively large for monthly data. This may hide other peaks and mislead the analysis. Besides, authors don't report the exact frequency of the peak, and we can't argue the period is one year without accurate calculation.

In "SARIMA" section, model selection is not reasonable. We list some major errors here: (1) The ACF / PACF plots are incorrect. Their ACF and PACF have lags "0.2, 0.4, 0.6, ...", which questions the reliability of these two plots. The analysis of these two plot is also problematic. For instance, their analysis of PACF provides no useful information. (2) Some statements are wrong. For instance, they claim "in the following, we conduct SARIMA model with previously selected components (p,d,q) from ARIMA and assume a seasonal period of 12", but we can't find their "previously selected components (p,d,q) from ARIMA". (3) They don't explain why they restrict (P, D, Q) to be $(0, 1, 0), (1, 1, 0), (0, 1, 1), (1, 1, 1)$ and why they finally choose $(P, D, Q) = (1, 1, 1)$. (4) They merely display the model roots in complex numbers and argue that the roots are all outside the unit circle. How can we know the magnitudes of these complex numbers? (5) The residual plot loses the values before 2008. (6) Four out of ten p-values of Ljung-Box test approximate 0.05, but authors say "most of the p-values seem above 0.05". They should further check the residuals. (7) Where is the comparison of predicted data and true data? We can only see the predicted data in their forecast plot. (8) **Conclusion is missing**.

In "References" section, references are poorly written. The references don't have index, and we can't find reference in previous text. The references are informal and not detailed (e.g., "R Markdown documentation", and "data source").

1.3 Points for Improvement

(1) Authors can conduct experiments on a lager and more representative dataset. (2) Authors' English writing needs to be improved. (3) Authors should focus on explaining the plots rather than just piling up plots. (4) Authors should arrange the contents of report more carefully.

2 Review: Project 5

2.1 Strength

The authors conduct a comprehensive time series analysis of the historical vehicle sales in the US, and their report has both conceptual and technical advantages. Conceptually, the topic is highly related to economy and US consumers, and it is worth delving into. The time series analysis successfully discovers the trend and seasonality within vehicle sales data. Their proposed SARMA model may provide useful guidance on the decision-making in the automotive industry. As for technical strength, their data analysis not only leverages multiple methods taught in class like periodogram and time series decomposition, but also includes some extra analyses such as PACF plot and yearly variance plot. Besides, the report is well written, and most contents are organized in an appropriate way.

2.2 Weakness

In "Exploratory Data Analysis" section, the analysis of median and mean is wrong. The reported mean and median are 1261.954 and 1268.462 respectively, but they argue that "the mean is higher than the median". The truth is that the mean is lower than the median. Besides, the skewness analysis may be also problematic. They state that "the data distribution is slightly right-skewed". However, if mean is less than median, we should get the analysis conclusion that the distribution is left-skewed, according to the rule of thumb "the mean is right of the median under right skew and left of median under left skew" [5]. Another minor issue is that we probably can't decide the skewness of distribution merely based on the comparison of mean and median. As pointed out by [5], the aforementioned rule of thumb frequently fails, especially when the distribution is multimodal or discrete. A more comprehensive analysis of the skewness may be needed.

In "Exploratory Data Analysis" section, the interpretation of PACF plot may be inaccurate. They claim that "the sharp cut-off after the first lag in the PACF suggests an AR(1) process might be appropriate for this time series". But we can find that the PACF plot actually suggests AR(3) model rather than AR(1). According to [4], "the partial autocorrelation for an AR(p) model is nonzero for lags $\leq$ p and 0 for lags $>$ p". When we look at their PACF plot, the partial autocorrelation decreases to 0 (i.e., PACF falls into the confidence interval) when $p > 3$, and this may imply that AR(3) model is more appropriate. However, strictly speaking, AR(3) model may be also inappropriate. This is because, for an AR(p) model, we should find the partial autocorrelations are all within confidence interval when lag is greater than p [4], but their PACF plot shows that there are many partial autocorrelations outside confidence interval even lag is greater than 3. In this case, we probably can't say AR model is suitable for this time series.

In "Stationarity Analysis" section, the ADF test is not correctly used and interpreted. According to [1]'s first comment, the ADF alternative hypothesis "the time series has no unit root i.e. it is stationary" is wrong, since the data may come from a non-stationary process that does not have a unit root. Besides, we can't say "the time series has no unit root" because the unit root is a property of models rather than data. The same problem also occurs in section "Stationary test after detrending".

In "Seasonal Subseries Plot" section, the seasonal subseries figure is not fully explained. At the beginning of this section, they state that the figure can identify the changes within particular seasons, but the subsequent text only analyzes the average sales for each month without discussing the changes within seasons. It seems that they leverage a very powerful ggplot function "gg_subseries" but don't fully understand its output.

In "Time Series Decomposition" section, the mechanism of STL is more complicated than what authors describe, and further explanation may be needed. For instance, the authors call the STL function without specifying the seasonal window size, and they claim that STL takes the default seasonal window size 13. However, when we look at the the documents of STL function [2,3], it shows that "the default (NULL) will choose an appropriate default, for a dataset with one seasonal pattern the seasonal window would be 11, the second larger seasonal window would be 15, then 19, 23, ... onwards", which contradicts authors' argument that the default seasonal window is hard-coded 13. Actually, the STL function will automatically tune its parameters according to the input data, and authors don't know the exact window size. When using these powerful black-box functions, authors should be more careful.

In "Spectral Analysis" section, authors ignore the existence of minor peaks. As shown in their raw periodogram, the peaks at frequency (roughly) 0.34 and 0.42 approximate the highest peak, and their confidence intervals overlap with that of highest peak. In this case, the authors probably can't directly argue that the period is 1 year without explaining why the minor peaks don't matter.

In "Modeling-ARIMA" section, some notations are confusing. For instance, they use a likelihood ratio test to compare two ARMA models: "a smaller model (arma1) and a larger model (arma2)", but don't specify the orders of these two models.

In "Modeling-ARIMA" section, their motivation of conducting likelihood ratio test (LRT) is not clear. It's not surprising to see LRT prefers ARMA(2, 2), considering that fact that ARMA(2, 2)'s AIC is much lower than its neighbor's. In this case, using LRT doesn't provide new insight about model selection.

In "Modeling-ARIMA" section, why we should select ARMA(2, 2) is not well justified: (1) The ACF plot may be better than Box-Ljung test in terms of identifying autocorrelation. Using ACF can clearly show at which lag the ACF value exceeds the confidence interval, and in this case, it can also possibly identify the seasonality. (2) Whether the residuals follow normal distribution is not tested. A QQ plot may be helpful. (3) The plot of AR and MA roots shows that the AR and MA roots almost cancel, which may suggest selecting a smaller model.

In "Modeling-SARIMA" section, the authors fix the (p, q) to be $(2, 2)$, $(1, 1)$ and $(1, 0)$ without justification. The authors pretend that the best SARMA models have $(p, q) = (2, 2)$, $(1, 1)$ or $(1, 0)$, and focus they study on searching (P, Q) values. However, this is counter-intuitive. In practice, the search range of (P, Q) is usually much smaller than that of (p, q). For instance, we usually test (p, q) ranging from 0 to 5, but only test (P, Q) of 0, 1 or 2. The large P or Q is prone to result in unstable model or convergence problem.

In "Modeling-SARIMA" section, the model selection is problematic. For instance, when $P \geq 1$ and $Q \geq 1$, the SARMA models have similar AIC values, and in this case, directly selecting the model of smallest AIC without testing its neighbors is not reasonable. Besides, they use Fisher standard error to test whether coefficients are significant, which is an error-prone method. Fisher standard error is valid only when some special conditions are satisfied, and usually, we should rely on bootstrap confidence interval rather than Fisher standard error.

In "References" section, the references are not in an appropriate format. Some references lack the author name. For instance, "[4] Seasonal Subseries Plot". Some references are just website links (e.g., the references [14], [15], [16]).

Some grammar errors and typos: (1) ("Exploratory Data Analysis" section) "The histogram of total vehicle sales displays a unimodal and xxx": we should use "an" instead of "a". (2) ("Seasonal Subseries Plot" section) "One possible explanation could be that people are less likely to making big purchases xxx": we should change "making" to "make".

2.3 Points for Improvement

(1) The algorithm of how to calculate PACF can be added for the ease of understanding. (2) More details about the mechanism of STL decomposition may be necessary. (3) It's interesting to see whether there is an association between vehicle sales and other factors like unemployment rate, GDP, and CPI. (4) A more

reasonable model selection strategy is needed.

References

[1] Anonymous Authors. The Comments on Project 05 (2022). https://ionides.github.io/531w22/midterm_project/project05/comments.html. Accessed: 2024-02-29.

[2] Rob J Hyndman and George Athanasopoulos. STL Decomposition. https://otexts.com/fpp3/stl.html. Accessed: 2024-02-29.

[3] Tidyverse. STL Decomposition. https://feasts.tidyverts.org/reference/STL.html. Accessed: 2024-02-29.

[4] Wikipedia. Partial Autocorrelation Function. https://en.wikipedia.org/wiki/Partial_autocorrelation_function. Accessed: 2024-02-29.

[5] Wikipedia. Skewness. https://en.wikipedia.org/wiki/Skewness. Accessed: 2024-02-29.