AAEC4484/AAEC(STAT)5484: Applied Economic Forecasting

Your Name Here

Homework #2 - Spring 2025

Instructions: Where necessary ensure that your graphs and visuals have proper titles and axis labels. Refer to the output, whenever appropriate, when discussing the results. **Creativity (coupled with relevance) will be rewarded.**

This week, the emphasis will be on pulling real-world data and conducting a more in-depth analysis of the data. The first question examines the Producer Price Index (PPI) series, while the second will explore the Weekly supply of Motor Gasoline in the US.

Question 1: Crude Oil Prices: West Texas Intermediate (WTI)

The WTI, a light (low density), sweet (low sulfur content) is a grade of crude oil used as one of the three global oil benchmarks. The other two are Brent Crude and Dubai Crude.

Although the Energy Information Administration (EIA) provides a wealth of data on crude oil prices, for this exercise, we will be using the `quantmod` package to import and analyze the monthly `WTI` series, `WTISPLC` from the Federal Reserve Bank of St. Louis (FRED).

a. Using the `getSymbols()` function from the `quantmod` package, import and produce an `autoplot` of the daily series, `DCOILWTICO`, from the FRED database (set `auto.assign = FALSE"`). For your convenience, I have already called in the `quantmod` library in the `setup` chunk.

Note: For this exercise, I do not require you to store the data. Instead, import the data and immediately plot the data and proceed to the next question.

b. Knowing that these are prices, is there anything that stands out as strange to you? Is there anything that explains the peculiarities you observe?

c. Given the peculiarities observed in the previous question, it would be best to work with the monthly series. To avoid any complicated calculations, we will return to the FRED database and import the monthly series, `WTISPLC`.

Proceed as follows:

- set `auto.assign = FALSE` and `return.class = "timeSeries"`
- Convert the data to a `ts` object using the `as.ts()` function.
- Convert the data to a `tsibble` object using the `as_tsibble()` function. This will allow us to use the time series functionalities of our `fpp3` package.
- Using the `filter_index()` function, subset the data to include only observations from `January 1990` to `December 2024`.
- **Store the results of this step into a variable called `wti`.**

d. Produce an `autoplot()`, and `ACF` plot of the `wti` series. Be sure to set the maximum lags in the ACF to be at least three (3) years. Comment on your plots. In particular, do you observe any patterns in the data? Does it appear to be a white noise process?

Use the `patchwork` package to combine your graphs. Also, use the `plot_annotation()` function to add a main title to your combined graph.

e. Compute the **logarithmic** growth rate of the `wti` series. For clarity, I am asking you to compute the first order logged difference. Be sure to multiply by 100 to get the percentage change. **It is fine to store this back in as `wti`.**

f. Repeat the analysis and interpretation of the plots in (d) for the `growth` series.

g. Anne and Vivian are study partners. From her read of the `ACF` plot, Anne is concerned about the initial lags and suggest that the growth series might actually not be a white noise process. Instead of arguing, Vivian opted to use the Ljung-Box test to test for autocorrelation up to lag $\ell = m \times 2$, $Q(24)$. What conclusions does Vivian draw at the *1%* level of significance? **Remember to state the null hypothesis and the decision rule of the test.**

Question 2: US finished motor gasoline products supplied

The US Energy Information Administration (EIA) reports several data series on the gas and renewable energy sector. For this exercise, we will be using the `Natural Gas Delivered to Consumers in the U.S. (Million Cubic Feet, MMcf)` series. The most updated data are available here

Our focus will be on estimating and evaluating competing models (albeit the most basic). At the end, you will select an "optimal" model for modeling the future value of monthly deliveries. This assignment should bridge the gap between the earlier modules and this current one.

Note: Before you proceed, please install the `rio` package. Do this in your consoles and not within the `.Rmd` file.

a. Using my code below, import the data from the EIA directly into R. If you were to download the `.xls` file manually, you would notice that the data is in the second sheet and that the first two rows are not part of the column headings you need. We will specify this using the `sheet = 2` and `skip = 2` arguments, respectively.

Please feel free to manually download the Excel file from the EIA link above and explore this on your own.

```
#remove eval=FALSE to make your codes work
gas.raw <- rio::import("https://www.eia.gov/dnav/ng/hist_xls/N3060US2m.xls", sheet = xxxxx,
                       skip = xxxx)
```

b. Next, let us clean up the data a bit. Your tasks are as follows:

- Using the `mutate()` function, declare the `Date` column as a `yearmonth()` object. It is okay to override `Date`.
- Rename the second column to `Delivery`. **Instead of typing out the full (and long) name of this column, a cool trick is to rename columns based on their name patterns. I would like you to use the `starts_with()` function to rename the second column.** A quick Google search will help you with this.
- Declare the dataset as a `tsibble` object using the `as_tsibble()` function.
- At the time of writing, the latest data was Dec, 2024. In the event the sheet has been updated by the time you attempt this assignment, use the `filter_index()` to ensure your sample period ends on `December 31, 2024`

Store the results of this step into `gas`.

c. Present an `autoplot`, `ACF`, and `gg_subseries` plot of the natural gas delivery series. Briefly comment on your plots. Set the maximum lags in the ACF to be at least three (3) years. Use `wrap_plots()` to present your graphs as a 3x1 grid. **Be sure to label your axes properly.**

d. Let us proceed with forecasting the NG delivery series. To do this, we will first need to split the data into a training (`train.gas`) and testing set (`test.gas`).

Assign the last two years of data to `test.gas` and the remaining observations to `train.gas`. **Note: Although we could use the `filter_index()` function to split the data, I would like you to use the `slice_head()` and `slice_tail()` functions instead.**

e. Confirm that the data is correctly split by using the `autoplot()` and `autolayer()` functions. Be sure to include `gas.ts`, `train.gas` and `test.gas` in this plot. **A title is not necessary. However, please add colors to your lines to highlight each series**.

f. Use the `model()` and `forecast()` functions to produce forecasts from the four(4) benchmark models from class. **Ensure that your models have appropriate names.**

- You will find it best to store both the model fits and forecasts for later.
- **Remember that you will need to set the horizon equal to the number of rows in your test set.**

g. Produce an `autoplot()` of the forecasts along with the *test* dataset. Be sure to turn off the `PI` at this stage. Does any particular forecast method appear to do a better job than others? Does any model appear to consistently over- or under-predict?

h. Using the `accuracy()` command, extract the `RMSE`, `MAE`, `ME`, and `MAPE` statistics. Also, I am intrigued to know the `MSE` value. Use the `mutate()` function to create that column.

- Display your results as a table using the `knitr::kable( )` function with `digits = 2`. Add the argument `format.arg = list(big.mark = ",")` to format your table using thousands separator (,000).
- Lastly, include **an appropriate caption** by adding and modifying the argument `caption = "Your caption Here"`.
- Comment on the bias of each model. **Be sure to also state which model in absolute terms is "least" biased.**
- **Ignoring the `ME`, which model is preferred under the remaining four (4) selection criteria? Be sure to explain how you came to that conclusion.**

i. As we discussed in class, the "preferred" model might not necessarily satisfy some or all of the necessary assumptions. Use the `gg_tsresiduals()` function, with `type = "innov"`, to comment on the **in-sample fit** of the model selected by the `MAPE`. **Again, set the maximum lags to 3 years so that you can see more of the dynamics of the `ACF`.**

When commenting, consider the following:

- Do the residuals appear to be normally distributed?

- Do you observe any **potential** patterns (seasonality, trend, cycle) in the `autoplot` and `ACF` of the residuals? From the patterns observed, do you think the residuals are white noise? Why, or why not?

j. Using the same model as above, employ the `Ljung-Box` test to conduct a hypothesis test of autocorrelation up to lag $\ell = 2 \times m$. **Where m is the frequency of the data.**

- You are required to explicitly state the null hypothesis of the test and how you came to your conclusion. **Think back to the decision rule of the test.**