Econometrics 475 Spring 2024
Homework #5 Due in Canvas Friday, March 22nd

Download the Excel dataset "HW5 Data.xlsx" from the class website. The "Crude" tab has a monthly data series over the period Jan. 2010 – Dec. 2023, for Texas Crude Oil Production in thousands of barrels.

1. Provide a time-series plot of the overall series.
2. Create a "training" data set ending June 2023:
    a. Use the ndiffs() function on the level of the series and report the order of differencing necessary to get a stationary series
    b. Use the ndiffs() function on the log of the series and report the order of differencing necessary to get a stationary series
    c. Produce plots of the appropriate differenced level and differenced log series; does a log transformation appear to be appropriate to stabilize the variance?
3. Use auto.arima() to allow R to choose the best ARIMA model for the series, *assuming the series is non-seasonal* (use seasonal=FALSE in auto.arima). If you determine that a log transformation is necessary to stabilize the variance, include the option lambda=0 in the auto.arima() statement.
    a. Report the summary of the fitted model.
    b. Use checkresiduals() for the fitted model; do they appear to be white noise? Provide the plots from checkresiduals() in your output along with your discussion.
    c. Use forecast() to generate forecasts through December 2023 (h=6) and provide a plot of the full series, the fitted values from auto.arima, and the forecasts.
    d. Report the forecast accuracy measures for the test set of July 2023 – December 2023.
4. Repeat part 3 allowing auto.arima() to determine if there are any seasonal elements in the series: use seasonal=TRUE in auto.arima(). How do the forecast accuracy measures compare to the results from part 3?

The "Elec" tab has a monthly data series over the period Jan. 2010 – Feb. 2024, for U.S. retail sales of Electronics and Appliance Stores in millions of dollars.

5. Provide a time-series plot of the overall series.
6. Create a "training" data set ending August 2023:
    a. Use the nsdiffs() function on the level of the series and report the order of seasonal differencing necessary to get a stationary series. If seasonal differencing is necessary, next use ndiffs() on the seasonally differenced series to check if additional differencing is needed. Report these results.
    b. Repeat part a on the log of the series.
    c. Produce plots of the appropriate differenced level and differenced log series; does a log transformation appear to be appropriate to stabilize the variance?
7. Use auto.arima() to allow R to choose the best ARIMA model for the series allowing seasonal effects (use seasonal=TRUE in auto.arima). If you determine that a log transformation is necessary to stabilize the variance, include the option lambda=0 in the auto.arima() statement.

d. Report the summary of the fitted model.
e. Use checkresiduals() for the fitted model; do they appear to be white noise? Provide the plots from checkresiduals() in your output along with your discussion.
f. Use forecast() to generate forecasts through February 2024 (h=6) and provide a plot of the full series, the fitted values from auto.arima, and the forecasts.
g. Report the forecast accuracy measures for the test set of September 2023 – February 2024.

Organize your results in PDF format to upload to Canvas along with a copy of your code.