

Assignment #4

Submit your **printed** solution at the start of class on Thursday, November 7.

Short answers, please. *Explain your answer concisely. If you refer to a plot in your answer, include that plot as part of your answer.* Do not include extraneous plots that you do not refer to in your narrative. “Significance” implies statistical significance. Presume necessary conditions for inference hold unless the question addresses these.

The time series in this assignment are weekly retail prices in dollars per gallon of regular conventional motor gasoline in the US. Questions 1-9 concern prices in 2023 and 2024 ($n = 90$); Question 10 considers prices in a longer time series, 2020-2024 ($n = 246$). The dates identify each week by the date of the Monday of that week.

Read the data from the file “a4_mg_price_23-24.csv” and use these to define weekly dates and prices. Then analyze the time series of prices.

1. Show a sequence plot of the weekly prices versus the weekly dates. The labels on the x-axis should be dates, not numbers of the weeks. Does the time series of the prices appear stationary?
[Because R’s time series do not handle weekly data naturally, you will need to generate your own sequence plot rather than rely on `tsplot` for this task. You might run into problems when combining a time series object with the built-in `plot` function. For some tasks, it’s easier to have a time series, and for others, a vector. The `lubridate` function `ymd` is very helpful with dates formatted like these.]
2. (a) Use the information in the ACF and PACF of the price time series to argue that an AR(2) model is appropriate for these data.
(b) Confirm the choice of an AR(2) model using the BIC criterion.
3. Fit an AR(2) model using maximum likelihood. Compare the estimates and standard errors produced by ML to those from a least squares regression. [`dynlm` requires a time series object, such as `price_t <- ts(price, start=1)`.]
4. Do diagnostics suggest inadequacies with the fitted AR(2) model computed in the prior question?
5. Generate forecasts for the next 36 weeks using the fitted model and show a graphical summary of these forecasts, along with their standard errors. [Like `dynlm`, `sarima.for` demands you input a time series to forecast. See Q3.]
6. Reproduce the forecasts for the first 3 weeks outside the observed data. That is, show how the R function `sarima.for` computes the forecast for the weeks of 2024-09-23, 2024-09-30, and 2024-10-07. Give detailed calculations for each value. You can (and should) use R as a calculator.
7. As in the prior question, show how the R function `sarima.for` computes the standard errors of the forecasts for the weeks of 2024-09-23, 2024-09-30, and 2024-10-07. As in Q6, give detailed calculations for each value.
8. The prices of motor gasoline in the 5 weeks following the period used to fit this model are \$3.074, \$3.087, \$3.026, \$3.069, and \$3.044. How well do forecasts from this model anticipate these values?

9. Use the AIC method to determine an alternative ARMA model for this time series. Compare the forecasts produced by the alternative model to those from the AR(2) model (for leads from 1 to 36 weeks out).

For the following question, use the time series in the file `a4_mg_price_20-24.csv`. The file is formatted as previously, but with more rows.

10. Show a sequence plot of the weekly prices on the weekly dates. Does this longer series appear stationary? Use additional plots or other statistics of your choosing as part of your answer.