Homework 5 – Due Tuesday, Mar. 21

STAT-GB.2302, STAT-UB.0018: Forecasting Time Series Data

The file `unemployment.csv` (original source: `https://research.stlouisfed.org/fred2/series/UNRATE/downloaddata`) contains the U.S. Civilian Unemployment Rate, monthly, from December, 1948 to the Present. The series has been seasonally adjusted.

(a) For the log of the unemployment series, use time series plots, ACFs, and PACFs to choose d for an ARIMA model. Explain your reasoning.

(b) With your choice of d from part (a), consider all 18 possible models with $p \leq 2$ and $q \leq 2$, with or without a constant. Use the AIC_C to identify p, q, and whether to include a constant in the model.

(c) Estimate the parameters. Are they all statistically significant? Note: R will not compute the p-values for you, but you can compute them yourself from the z statistics (which you can compute yourself by taking the parameter estimate and divided by the standard error).

(d) Write the complete form of the fitted model. (For example, $x_t = .3x_{t-1} + .2x_{t-2} + \varepsilon_t + .4\varepsilon_{t-1} - 2.351$.)

(e) Examine the Ljung-Box statistics for lack of fit for the first 12, 24, 36, and 48 lags. According to these statistics, does the model seem to be adequate? The model is declared to be inadequate if the p-value is less than $.05$ (in this case, we reject the null hypothesis that the model is adequate).

(f) Plot the residuals from the fitted model, as well as the ACF and PACF of the residuals. Do these plots indicate any inadequacies in the model?

(g) Obtain forecasts and 95% forecast intervals for lead times 1 to 30. Plot the data, forecasts, and forecast intervals on a single plot. Do the forecasts seem reasonable? Do the forecast intervals seem excessively wide?

New R commands used in this assignment

To complete this assignment, you will need to use the following new R commands.

- `Arima`. Fit an ARIMA model. For models without a constant, the model is of the form

  ```
  x(t) = ar1 * x(t-1) + ar2 * x(t-2) + ... + arp * x(t-p)
       + eps(t) + ma1 * eps(t-1)  + ma2 * eps(t-2) + ... + maq * eps(t-q)
  ```

  where `x(t)` is the series, and `eps(t)` is a white noise. For models with a constant, the model is of the form

  ```
  [x(t) - C] = ar1 * [x(t-1) - C] + ar2 * [x(t-2) - C] + ... + arp * [x(t-p) - C]
             + eps(t) + ma1 * eps(t-1)  + ma2 * eps(t-2) + ... + maq * eps(t-q)
  ```

  When $d = 1$, R refers to the constant, `C` as an "intercept"; when $d = 2$, R refers to `C` as a "drift".

- `Box.test`. Perform a Ljung-Box or Box-Pierce test for examining the null hypothesis of independence in a given time series. You must specify `lag`, the number of lags to consider, along with `fitdf`, the degrees of freedom in the fit. For an ARIMA(p, d, q) without a constant, `fitdf` should be $p + q$; with a constant, `fitdf` should be $p + q + 1$. Examples:

  ```
  # 12 lags, Model without constant:
  fit1 <- Arima(x, c(p, d, q), include.constant=FALSE)
  resid1 <- residuals(fit1)
  Box.test(resid1, lag = 12, type = "Ljung-Box", fitdf = p + q)

  # 24 lags, Model with constant:
  fit2 <- Arima(x, c(p, d, q), include.constant=TRUE)
  resid2 <- residuals(fit2)
  Box.test(resid1, lag = 24, type = "Ljung-Box", fitdf = p + q + 1)
  ```

- `forecast`. Produce forecasts and forecast intervals from a fitted time series model. Examples:

  ```
  fit <- Arima(x, c(p, d, q), include.constant=TRUE)

  # Compute forecasts for lead times 1 to 20, with 95% PI
  forecast(fit, h = 20, level = 95)

  # Plot forecasts for lead times 1 to 10, with 99% PI
  plot(forecast(fit, h = 10, level = 99))
  ```