STAT 520, Fall 2015: Homework 5

1. In Homework 4, you were asked to identify a small set of candidate ARIMA(p, d, q) models for each of the following data sets:
   - `ibm`: daily closing IBM stock prices (dates not given)
   - `thermal`: daily temperatures data used in midterm
   - `gasprices`: average price (US dollars per gallon) for regular gasoline in the United States; there are $n = 145$ weekly observations collected from 1/5/2009 to 10/10/2011.
   - `supremecourt`: data used in Example 1.15 on p. 16 of the notes.

   (a) After a potential transformation, fit each set of candidate models (3 or fewer models) for each data set, using maximum likelihood (the default) and then perform the diagnostic checks outlined in Chapter 8: check residuals for normality (histogram, qq-plot, Shapiro-Wilk test), for independence (runs test), for outliers, the ACF, and the Ljung-Box test. The last three are automatic in `tsdiag`.

   (b) Choose a final model (if all fit fine, choose the one with the smallest AIC) for each data set and calculate MMSE forecasts and prediction intervals for future values. You decide "how far out" in time to forecast. Display the forecasts and prediction bands visually like in the notes; sample R code for Chapter 9 is available on the web page.

   (c) For each data set, perform overfitting for the final model and comment. Also write down your final fitted model, as in the notes.

   (d) Compare your forecasts to those obtained via the `auto.arima` function in the `forecast` R package. Include `lambda=F` if there is no initial transformation of the data. If, e.g. a natural log transformation is warranted, include `lambda=0`. Sample code:

```
library(forecast)
f=auto.arima(d1,lambda=0) # log-transformation of original data
plot(forecast(f,h=20)) # includes all data
plot(forecast(f,h=20),include=50) # only include last 50 observations
```

   Note that the two shaded bands correspond to 80% and 95% prediction intervals. How do the `auto.arima` forecasts compare to yours?