Assignment #3

Remark: The goal of this assignment is to provide students some opportunity to analyze financial time series from different markets. Real financial time series used in this assignment contain possible outliers. But those outlying data points might be due to jumps and time-varying volatility, two common features of financial data. We shall discuss volatility modeling and heavy-tail distributions later in class. The number of outliers should reduce substantially after that.

1. Consider the monthly new one family houses sold in the U.S. from January 1963 to February 2017. The data are available from FRED and are entitled `HSN1FNSA`.

   (a) Build an Airline model for the house-sold data. You should perform model checking. Write down the fitted model.

   (b) Denote the series by `hsold`. Fit the following model
   ```
   m2 <- arima(hsold, order=c(0,1,1),seasonal=list(order=c(1,1,1),period=12))
   ```
   Are all coefficient estimates significant at the 5% level? If not, use the subcommand `fixed` to remove insignificant estimates. Write down the fitted model.

   (c) Compare the models in parts (1) and (2). Which model is preferred in the in-sample comparison? Why?

   (d) Use the `backtest` to compare the two models with forecast origin $t = 600$. Which model do you prefer? Why?

2. Again, consider the logarithm of daily VIX index. The time span is from January 1999 to March 31, 2017. You may use the following command from `quantmod` to down load the data

   ```
   require(quantmod)
   getSymbols(``^VIX'',from=``1999-01-03'',to=``2017-03-31'')
   vix <- log(as.numeric(VIX[,6]))
   ```

   (a) Build a time series model for the log VIX index, including model checking.

   (b) Write down the fitted model.

(c) Obtain 1-step to 5-step ahead point predictions of the log VIX index at the forecast origin March 31, 2017.

(d) Refine the model by handling the largest (in absolute value) three outliers.

3. Consider the monthly 10-year treasury constant maturity rate and 1-year treasury constant maturity rate from April, 1953 to March 2017. The two series can be downloaded from "FRED". The names are `GS10` and `GS1`, respectively. Denote the 10-year rate as y_t and the 1-year rate as x_t.

   (a) Fit the simple linear regression model **(without time series errors)**

   $$y_t = \alpha + \beta x_t + e_t.$$

   Write down the fitted model, including residual standard error and R^2. Is the model adequate? Why?

   (b) Build a regression model for y_t using x_t as the explanatory variable, including model checking. **(This should be a regression with time series errors.)**

   (c) Write down the fitted model of part (b).

4. Consider the daily CDS spreads (3-year maturity) of Allstate Insurance from January 01, 2004 to September 19, 2014. The period includes the financial crisis of 2008 so that the CDS spreads vary substantially. The data are in the file `d-cdsALL.txt` (column 2). Since the spreads are small, we consider the time series $x_t = 100 \times$ (spread3y). In addition, sample ACF of x_t shows strong serial dependence. Therefore, we analyze the differenced series $y_t = (1 - B)x_t$.

   (a) Build a time series model for y_t. Write down the fitted model. [You may start with a model suggested by the `auto.arima` command in `forecast` package.]

   (b) Is the model obtained in part (a) adequate? Why?

   (c) To improve the fit, identify sequentially the largest four outliers of the fitted model. Write down the fitted model with the four largest outliers included.

   (d) Let a_t be the residuals of the model in part (c) and ρ_i be the lag-i ACF of a_t. Test $H_0 : \rho_1 = \cdots = \rho_{10} = 0$ versus $H_a : \rho_i \neq 0$ for some $1 \leq i \leq 10$. Draw your conclusion.

5. Consider, again, the daily CDS spreads of Allstate in Problem 4. Now, consider $z_t = \log(x_t)$. Let $d_t = (1 - B)z_t$.

   (a) Build a time series model for d_t. Write down the fitted model.

   (b) Is the model obtained in part (a) adequate? Why?

   (c) To improve the fit, identify sequentially the largest two outliers of the fitted model. Write down the fitted model with the two largest outliers included.

(d) Let a_t be the residuals of the model in part (c) and ρ_i be the lag-i ACF of a_t. Test $H_0 : \rho_1 = \cdots = \rho_{10} = 0$ versus $H_a : \rho_i \neq 0$ for some $1 \leq i \leq 10$. Draw your conclusion.

(e) Comment on the effect of the log transformation by comparing models of Problems 4 and 5.