## STAT 6351: Assignment #3

## Notes

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.

## Questions

- 1. Consider the monthly new single-family homes sold in the US from January 1963 to February 2017. The data are available from FRED and are entitled HSN1FNSA. Denote the series by hsold.
  - (a) Build an Airline model for hsold. You should perform model checking. Write down the fitted model (call it m1).
  - (b) Fit the following model:
    - m2 <- arima(hsold, order=c(0,1,1),seasonal=list(order=c(1,1,1),period=12))</pre>

Are all coefficient estimates significant at the 5% level? If not, use the option fixed to remove insignificant estimates.

- (c) Compare the models m1 and m2. Which model is preferred in the in-sample comparison? Justify your answer.
- (d) Use the function backtest to compare the two models with forecast origin t = 600. Which model do you prefer? Justify your answer.

2. Consider again the log of daily VIX index. The time span is from January 1999 to March 31, 2017. You may use the following command from quantmod to download the data:

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

- (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 3 outliers (in absolute value).
- 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):

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

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.) Write down the fitted model. Is this more refined model adequate? Why?
- 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 2nd column of the file d-cdsALL.txt, and is called spread3y. Since the spreads are small, we consider the series  $x_t = 100 \times (\text{spread3y})$ . Since the sample ACF of  $x_t$  shows strong serial dependence, 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 function in the forecast package.)
  - (b) Is the model obtained in (a) adequate? Why?
  - (c) To improve the fit, identify sequentially the largest 4 outliers of the fitted model. Write down the fitted model with these largest outliers included.
  - (d) Let  $a_t$  be the residuals of the model in (c), and  $\rho_h$  be the lag-*h* ACF of  $a_t$ . Test  $H_0: \rho_1 = \cdots = \rho_{10} = 0$  vs.  $H_a: \rho_h \neq 0$ , for some  $1 \leq h \leq 10$ . Draw your conclusion.

- 5. Consider again the daily CDS spreads of Allstate in Problem 4. Now, let  $z_t = \log(x_t)$  and  $d_t = (1 B)z_t$ .
  - (a) Build a time series model for  $d_t$ . Write down the fitted model. Is the model adequate? Why?
  - (b) To improve the fit, identify sequentially the largest 2 outliers of the fitted model. Write down the fitted model with these largest outliers included.
  - (c) Let  $a_t$  be the residuals of the model in (b), and  $\rho_h$  be the lag-*h* ACF of  $a_t$ . Test  $H_0: \rho_1 = \cdots = \rho_{10} = 0$  vs.  $H_a: \rho_h \neq 0$ , for some  $1 \leq h \leq 10$ . Draw your conclusion.
  - (d) Comment on the effect of the log transformation by comparing the models of Problems 4 and 5.