Homework 7

Due: Thursday 4/18/19 by 12:00pm (noon)

Note - problems 2. and 3. will require use of the the `MARSS` package for `R`.

1. Exploratory and ARIMA Analysis of Project Data

For this problem, I'll ask you to select a data set from the following possibilities:

- Anomaly
- Stocks
- Yields

(a) Give the name of the dataset you've chosen. You'll have to stick with this dataset for the ARIMA part of the final project. All but one of the datasets are multivariate. For this problem, I will ask you to analyze the first time series in the dataset you've selected. For instance, the first time series in the `Anomaly` data is given by `Anomaly[, 1]`.

(b) Make a plot with three panels using `par(mfrow = c(1, 3))`. In the first panel, plot the raw data. In the second panel, plot the first differenced data. In the third panel, plot the second differences.

(c) Using the `ndiffs` function, determine the amount of differencing needed based on a level-α augmented Dickey-Fuller test. Perform the tests with a time trend included.

(d) Based on (b) and (c), how many times do you think you should difference the data if you plan to use an $\mathbf{ARIMA}(p, q)$ model? Provide a reason in at most one sentence.

(e) Make a plot with two panels using `par(mfrow = c(1, 2))`. In the first panel, plot the sample autocorrelations up to lag $l = 15$ for the differenced data using the amount of differencing chosen in (d). In the second panel, plot the sample partial autocorrelations up to lag $l = 15$ for the differenced data using the amount of differencing chosen in (d). Use the `acf` function for both.

(f) Based on the results of (e), indicate the number of AR terms you would include if you were to fit an $\mathbf{ARIMA}(p, d, 0)$ model, and the number of MA terms you would include if you were to fit an $\mathbf{ARIMA}(0, d, q)$ model.

(g) Find the best $\mathbf{ARIMA}(p, d, q)$ model using the differencing parameter d chosen in (d) using AIC, using maximum values and p and q that correspond to your ACF and PACF-based values for pure **MA** and **AR** models in (f). Make sure you allow for a nonzero intercept in the model for the differenced time series. If the `arima` function does not converge for certain values of p and q, try setting method equal to `ML`. If you are getting errors that stop your for loop or indicate that things didn't converge even though an `AIC` value is returned, you can deal with them by saving your `fit` (what I would call my fitted **ARIMA** object) using the `try` function, which won't break your code but will return `TRUE` if you type `inherits(fit, "try-error")`. You can also identify non-convergence by checking if `!fit$code == 0`. If there was an error that prevented `arima` from running or the model did not converge, just set the AIC for that model to `NA`. Here is an example:

```r
set.seed(1)
x <- rnorm(100)
fit <- try(arima(x, order = c(30, 0, 30))) # This will give you an error
if (inherits(fit, "try-error") | !fit$code == 0) {
  aic <- NA
} else {
```

```r
  aic <- AIC(fit)
}
```

(h) Plot the 10-step ahead forecasts and corresponding 95% intervals based the $\mathbf{ARIMA}(p, d, 0)$, $\mathbf{ARIMA}(0, d, q)$, and $\mathbf{ARIMA}(p, d, q)$ models from (f) and (g) on the same plot, on the same scale as the original data. Include the raw time series on the plot as well.

(i) In at most two sentences, indicate which of the models in (h) you prefer and explain why.

2. Understanding State Space Models

In this probem, we're going to play around with the parameters of the state-space model a bit, building off of the example of the `varve` data we saw in class. The `varve` data is available in the `astsa` package. We'll work with it on the log scale, because the variance of the observations appears to be increasing over time. In class, we fit the following model to the data:

$$y_t = ax_t + v_t \text{ and } x_t = \phi x_{t-1} + w_t,$$

where $v_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_v^2\right)$, $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_w^2\right)$, and $x_0 = \mu$.

```r
rm(list = ls())

library(astsa)
library(MARSS)

data("varve")
n <- length(varve)

y <- log(varve)
plot(y)

model <- list(
  B=matrix("phi"), U=matrix(0), Q=matrix("sig.sq.w"),
  Z=matrix("a"), A=matrix(0), R=matrix("sig.sq.v"),
  x0=matrix("mu"), tinitx=0 )

fit <- MARSS(c(y), model=model, method = "kem")
# Get Kalman filter, predictor and smoother
kf <- MARSSkfss(fit)
```

The estimated coefficients were $\hat{a} = 0.9776$, $\hat{\sigma}_v^2 = 0.1805$, $\hat{\phi} = 0.999$, $\hat{\sigma}_w^2 = 0.0133$, and $\hat{\mu} = 3.5267$.

For this problem, I would like you to plot the predictions, filter, and smoother for the following state-space models:

i. Keeping $\hat{a}$, $\hat{\phi}$, $\hat{\sigma}_w^2$, and $\hat{\mu}$ at their estimated values and setting $\sigma_v^2 = 0.05$.
ii. Keeping $\hat{a}$, $\hat{\phi}$, $\hat{\sigma}_w^2$, and $\hat{\mu}$ at their estimated values and setting $\sigma_v^2 = 0.5$. iii.Keeping $\hat{a}$, $\hat{\phi}$, $\hat{\sigma}_v^2$, and $\hat{\mu}$ at their estimated values and setting $\sigma_w^2 = 0.0001$.
iii. Keeping $\hat{a}$, $\hat{\phi}$, $\hat{\sigma}_v^2$, and $\hat{\mu}$ at their estimated values and setting $\sigma_w^2 = 1$.

(a) Make a plot with twelve panels using `par(mfrow = c(4, 3))`. In each panel, plot the first 50 observations from the data in gray.

- In the first column of panels, add solid black lines for the predictions of the states using the estimated parameters and dashed black lines for the predictions of the states using the estimated parameters plus and minus the corresponding standard errors.

- In the second column of panels, add solid black lines for the filtered values of the states using the estimated parameters and dashed black lines for the filtered values of the states using the estimated parameters plus and minus the corresponding standard errors.
- In the third column of panels, add solid black lines for the smoothed values of the states using the estimated parameters and dashed black lines for the smoothed values of the states using the estimated parameters plus and minus the corresponding standard errors.
- In the first row of panels, add lines for the the predictions, filtered values, ad smoothed values from model i.
- In the second row of panels, add lines for the the predictions, filtered values, ad smoothed values from model ii.
- In the third row of panels, add lines for the the predictions, filtered values, ad smoothed values from model iii.
- In the fourth row of panels, add lines for the the predictions, filtered values, ad smoothed values from model iv.

(b) In at one sentence, explain how the value of σ^2_v affects affects the conditional means and variances of the states given the observed data.

(c) In at one sentence, explain how the value of σ^2_w affects the conditional means and variances of the states given the observed data.

3. A State Space Model for the Chicken Data

In this problem, we're going to further examine the use of the state-space model

$$y_t = ax_t + v_t \text{ and } x_t = \phi x_{t-1} + w_t,$$

where $v_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2_v\right)$, $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2_w\right)$, and $x_0 = \mu$.

We will use the `chicken` data from the `astsa` package once more. A special case of this state-space model is the $\mathbf{ARIMA}(1, 0, 0)$ model.

(a) Fit the state space model to the `chicken` data with $a = 1$ and $\sigma^2_v = 0$, as well as an $\mathbf{ARIMA}(1, 0, 0)$ model. You will need to use `method=ML` when you run `arima` to fit this model. Compare the parameter estimates. In at most one sentence, indicate whether or not the estimates of the remaining parameters are the same and explain why or why not this is the case.

(b) Fit the entire state space model with unknown a and σ^2_v to the chicken data with the last 20 observations removed. On a plot with the x-axis ranging from $2001 + 7/12$ to $2016 + 6/12$ and y-axis randing from 60 to 130, plot:

- The observed chicken time series.
- Predicted chicken chicken prices from $2014 + 11/12$ to $2016 + 6/12$ based on the $\mathbf{ARIMA}(1, 1, 0)$ model.
- Predicted chicken prices from $2014 + 11/12$ to $2016 + 6/12$ based on the $\mathbf{ARIMA}(1, 1, 0)$ model plus or minus one standard error.
- Forecasted chicken prices from $2014 + 11/12$ to $2016 + 6/12$ based on the state-space model.
- Forecasted chicken prices from $2014 + 11/12$ to $2016 + 6/12$ based on the state-space model plus or minus one standard error.

(c) In at most one sentence, indicate which model you prefer for the chicken data and explain why.