Homework 8

Due: Wednesday 5/1/19 by 12:00pm (noon)

Note - problem 2. will require use of the the `stochvol` package for `R`.

1. Exploratory and State-Space Analysis of Project Data

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

- Anomaly
- Electricity
- Stocks
- Yields
- Air
- Beijing

(a) Give the name of the dataset you've chosen. You'll have to stick with this dataset for the state-space 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) Plot the raw data.

(c) All of the data sets have some type of "seasonal" aspect, i.e. they are measured quarterly, monthly, or daily and may have quarter-of-the-year, month-of-the-year, or day-of-the-week effects, respectively. What kind of seasonality might be present in the data you chose?

(d) Define an $n \times s$ matrix $\boldsymbol{Z}$ to capture seasonality, where s is the number of units of time per season minus one. Note that you'll have to transpose it when you pass it to the `MARSS` function(s). Fit four linear state-space models to the raw data minus the last 20 observations using the `MARSS` package:

- i. $y_t = ax_t + v_t$, $x_t = \phi x_{t-1} + w_t$
- ii. $y_t = ax_t + \boldsymbol{\gamma}'\boldsymbol{z}_t + v_t$, $x_t = \phi x_{t-1} + w_t$
- iii. $y_t = ax_t + v_t$, $x_t = \phi x_{t-1} + \boldsymbol{\upsilon}'\boldsymbol{z}_t + w_t$
- iv. $y_t = ax_t + \boldsymbol{\gamma}'\boldsymbol{z}_t + v_t$, $x_t = \phi x_{t-1} + \boldsymbol{\upsilon}'\boldsymbol{z}_t + w_t$

Compute AIC for each, and indicate which model you would choose based on AIC alone.

(e) Plot the last 40 observations from the raw data, the forecasts of the last 20 observations under each model, and 95% confidence intervals for each.

(f) Compute the average squared forecast error for the last 20 observations under each of the four models. Indicate which would you choose based on squared forecast error alone.

(g) In at most one sentence, indicate whether or not you would choose a model based on AIC or squared forecast error and explain why.

2. Stochastic Volatility

On the second exam, we applied a **GARCH**$(m, 0)$ model to the second differences of the demeaned `gnp` data. We're going to apply a stochastic volatility model to the same data. You'll want to start with the following code to load the packages we need and the data:

```r
library(astsa)
library(stochvol)
data(gnp)
y <- (diff(gnp, d = 2) - mean(diff(gnp, d = 2)))
n <- length(y)
```

(a) Fit stochastic volatility model with the default prior specifications to `y` using the `svsample` function **three times for each** of the following values of m, the number of simulated values of the states and parameters drawn from the posterior distribution:

   i. $m = 100$;
   ii. $m = 1000$;
   iii. $m = 10000$.

Make a plot with three panels, one for each value of m. In each panel, plot the estimates of the posterior means for the latent states $\boldsymbol{h}$ for each run of `svsample`. You will have three lines per panel.

(b) For this data, which value of m seems reasonable to use in practice? Answer in at most one sentence and base your answer on your plots from (a).

(c) Using the value of m you argued for in (b), fit the stochastic volatility models to the data with the last 20 observations held out using the following priors:

   i. Default priors for μ_h, ϕ and σ_w^2;
   ii. Default priors for ϕ and σ_w^2, normal prior for μ_h with mean 0 and variance 1;
   iii. Default priors ϕ and σ_w^2, normal prior for μ_h with mean 0 and variance 1000000;
   iv. Default priors for μ_h and σ_w^2, beta prior for $(\phi + 1)/2$ with $a_0 = 1$ and $b_0 = 1$;
   v. Default priors for μ_h and σ_w^2, beta prior for $(\phi + 1)/2$ with $a_0 = 10$ and $b_0 = 10$;
   vi. Default priors for μ_h and ϕ, gamma prior for σ_w^2 with shape 1/2 and rate 1/20.
   vii. Default priors for μ_h and ϕ, gamma prior for σ_w^2 with shape 1/2 and rate 1/0.02.

Plot kernel density estimates of $p(\mu_h|\boldsymbol{y})$ from i.-iii. in the first panel, kernel density estimates of $p(\phi|\boldsymbol{y})$ from i., iv-v. in the second panel, and $p\left(\sigma_w^2|\boldsymbol{y}\right)$ from i., vi-vii. in the last panel. Kernel density estimates can be computed using the `density` function in `R` applied to simulated values of the corresponding parameter returned by `svsample`.

(d) Give the average squared forecast error for the last 20 observations for all of the models fit in (c) in a table. For your forecasts, use the average simulated value of each future y_{n+k}, which can be obtained using the the `predict` function.

(e) Based on what you observe (c) and (d), explain in one sentence which prior specification(s) you prefer. You don't have to choose a single one, but you should comment on whether or not any seem like especially good or bad choices.