Stat 5170: Assignment 11

due May 4, 5pm

1 R Tutorial

1. In this example, we will explore how to fit a linear regression model with AR errors. The dataset for this first example is "varve.dat". The data measure the thickness of deposits of sand and silt (varve) left by spring melting of glaciers about 11800 years ago. The data are annual estimates of varve thickness at a location in Massachusetts for 455 years. Read the data in. Due to nonconstant variance of the variable, we will apply a $\log_{10}$ transformation.

```
y<-ts(scan("varve.dat"))
y.t<-log(y, base=10)
```

We seek to fit the following model

$$\log_{10} y = \beta_0 + \beta_1(t - \bar{t}) + \beta_2(t - \bar{t})^2 + \epsilon \tag{1}$$

where $t = 1, 2, \cdots, 455$ is the number of the year. To create these variables, type

```
year<-seq(1,455,1)
year.t<-year-mean(year)
year.t2<-year.t^2
```

(a) Fitting OLS regression. Use `result<-lm(y.t~year.t+year.t2)` to fit model (1). Use the `summary()` function to obtain the estimated coefficients, their standard errors, and the standard error for the model.

(b) Independence of residuals. For an OLS model to be valid, we require the residuals to be white noise. We can use the ACF and PACF plots of the residuals to determine if this condition is met. The `residuals()` function obtains the sample residuals of a model. Then plot the ACF and PACF for the residuals.

(c) Estimating the AR operator. Based on the ACF and PACF plots, the residuals appear to have an AR(1) structure, so we need an estimate for ϕ. Type `ar1<-arima(residuals(result),c(1,0,0))`. You may choose to add an optional argument `include.mean=F` in the `arima()` function.

(d) Transforming the variables. We want to apply the AR(1) operator to each of the variables. To do so, type

```
y.mat<-cbind(as.ts(y.t), lag(y.t,-1))
year.mat<-cbind(as.ts(year.t), lag(year.t,-1))
year.mat2<-cbind(as.ts(year.t2), lag(year.t2,-1))

y.new<-y.mat[,1] - ar1$coef*y.mat[,2]
year.new<-year.mat[,1] - ar1$coef*year.mat[,2]
year.new2<-year.mat2[,1] - ar1$coef*year.mat2[,2]
```

(e) Regress the transformed variables. Use `result2<-lm(y.new~year.new+year.new2)` and the `summary()` function.

(f) Model diagnostics. Check the independence of the residuals from the model with the transformed variables. They should appear to be white noise. If they don't, then the transformation you used might be wrong.

2 Assignment

1. The datasets "gas82.dat" and "oil82.dat" give gas price and oil price indices in the United States for 82 consecutive months. Well treat gas as the explanatory variable x_t and oil as the response variable y_t.

   (a) Fit the simple linear regression model using ordinary least squares, $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$. Report the estimated intercept and slope along with their standard errors.

   (b) Produce the ACF and PACF of the residuals from the regression in part 1a. Based on the plots, what AR model is appropriate for the residuals?

   (c) Estimate the coefficient(s) of the AR model you determined in part 1b. What transformation of y_t and x_t is appropriate to model them in a regression model with white noise errors?

   (d) Apply the transformation stated in part 1c to y_t and x_t. Fit the simple regression model to your transformed y_t and transformed x_t. Produce the ACF and PACF plots of the residuals from this model. Based on these plots, is this model appropriate?

   (e) Give the estimated coefficients, along with their standard errors, for the model in Part 1d. Compare these values to those in Part 1a. Comment on whether you trust the standard errors from the model in Part 1a and Part1d.