Stat 6021 R Tutorial: Regression with AR Errors

For this tutorial, we will use the data set "company.txt". The variables *company* and *industry* are the company sales in millions, and industry sales in millions respectively. The data are collected over 20 quarters. The company wishes to predict its sales by using industry sales as a predictor. Since the variables are time-dependent, ordinary least-squares (OLS) regression is likely to be inappropriate since the observations are likely to be correlated, which leads to errors that are autocorrelated. In this tutorial, you will learn how to fit a simple linear regression model with autoregressive (AR) errors.

As a reminder, the simple linear regression model with AR(p) errors is:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \tag{1}$$

where

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_p \epsilon_{t-p} + a_t, \tag{2}$$

and a_t are i.i.d $N(0, \sigma_a^2)$. The Cochrane-Orcutt method transforms the variables based on the AR structure of the errors in (2):

- $y_t' = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p}$
- $x_t' = x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p}$

Subbing (1) into these transformations, the model becomes

$$y_t' = \beta_0' + \beta_1' x_t' + a_t \tag{3}$$

so the errors from (3) are now i.i.d. $N(0, \sigma_a^2)$.

The steps in fitting a regression with AR errors include:

1. Use ordinary least squares to estimate the coefficients in (1).
2. Examine the AR structure of the sample residuals from step 1 using a PACF Plot.
3. Estimate the coefficients $\phi_1, \cdots, \phi_p$ using ARIMA estimation (`arima()` function).
4. Use the estimated $\hat{\phi}_1, \cdots, \hat{\phi}_p$ to compute y_t' and x_t'.

5. Use OLS to estimate (3) using y'_t and x'_t.
6. Check model assumptions using residual plot, ACF plot, and normal probability plot of residuals. If the transformation was done correctly, the residuals should have mean 0, constant variance and are uncorrelated.

So, we can implement steps 1 to 6 using the following:

1. Fit a least squares regression to the data as usual.

   ```
   result<-lm(company~industry)
   ```

2. Obtain a PACF plot of the residuals from the regression model.

   ```
   res<-result$residuals
   pacf(res, main="PACF of Residuals")
   ```

   Notice that the PACF plot starts at lag 1. The ACF plot created with the `acf()` function starts at lag 0.

3. Since the PACF is significant at lag 1, but not at any other lag, the residuals follow an AR(1) structure, so we only need to estimate ϕ_1, which can be done using the `arima()` function, i.e.,

   ```
   ar.1<-arima(res,order = c(1,0,0), include.mean = FALSE)
   ar.1$coef
   ```

   The argument order = c(1,0,0) indicates we want to fit an AR(1) structure to the error terms. What is the estimate for ϕ_1 here?

4. Since the errors are fitted with an AR(1) structure, use the following transformation for both variables: $y'_t = y_t - \hat{\phi}_1 y_{t-1}$ and $x'_t = x_t - \hat{\phi}_1 x_{t-1}$.

   ```
   shift<-ar.1$coef
   y<-cbind(as.ts(company),lag(company))
   yprime<-y[,2] - shift*y[,1]
   x<-cbind(as.ts(industry),lag(industry))
   xprime<-x[,2] - shift*x[,1]
   ```

5. Use least squares to fit the transformed variables.

   ```
   result.prime<-lm(yprime~xprime)
   summary(result.prime)
   ```

   How do we interpret the t test for β'_1?

6. As a final step, we should examine the residual plot, the ACF plot, and normal probability plot of the residuals from the regression model.

```
##residual plot
plot(result.prime$fitted.values,result.prime$residuals, main="Plot of Residuals aga
abline(h=0,col="red")

##acf plot of residuals
acf(result.prime$residuals)

##qq plot of residuals
qqnorm(result.prime$residuals)
qqline(result.prime$residuals, col="red")
```