Lab3

Kimberly Boswell

10/18/2023

Common Questions Received

1) Changing categorical variables to numeric Either use an ifelse statement or leave as a factor. You can check if it is a factor using class(df$variable). If not, you can use factor(variable) to convert it within the regression and as.numeric(factor) to convert it to a numeric variable.

Using `SwissLabor` in AER, evaluate the type of variables, regress income against all variables in the dataset and show results. Add a column which converts participation to an indicator variable with 1 for yes and 0 for no. Check its status.

2) "HH" not working for regsubsets output? Use forward or backward selections with `step` or `stepsAIC`

Example:

Can we use age and education to determine an individual's nonlabor income? Using `SwissLabor` in AER, let us inspect and select a set of variables using forward and backward selection.

```
data("SwissLabor")
swissreg <- lm(income ~ ., data = SwissLabor)
optreg <- step(swissreg, direction = "backward")

# optreg1 <- stepAIC(swissreg, direction = "backward")
# You need the MASS package for this

#or

null.model <- lm(income ~ 1, data = SwissLabor)
full.model <- lm(income ~ ., data = SwissLabor)
optreg2 <- step(lm(income ~ 1, data = SwissLabor), direction = "forward",
                scope=list(lower=null.model, upper=full.model) )
```

Time Series and Autocorrelation

Continuing from last week:

1) Load okun5_aus
2) Check that it is a time series
3) If not, make it a time-series
4) Observe the lags of the variable.
5) Regress the differenced unemployment rate against 02, 3 and 4 lags of growth.
6) Construct a timeseries plot and correlogram for growth rate and unemployment.

Converting Data Frame to Time Series Object

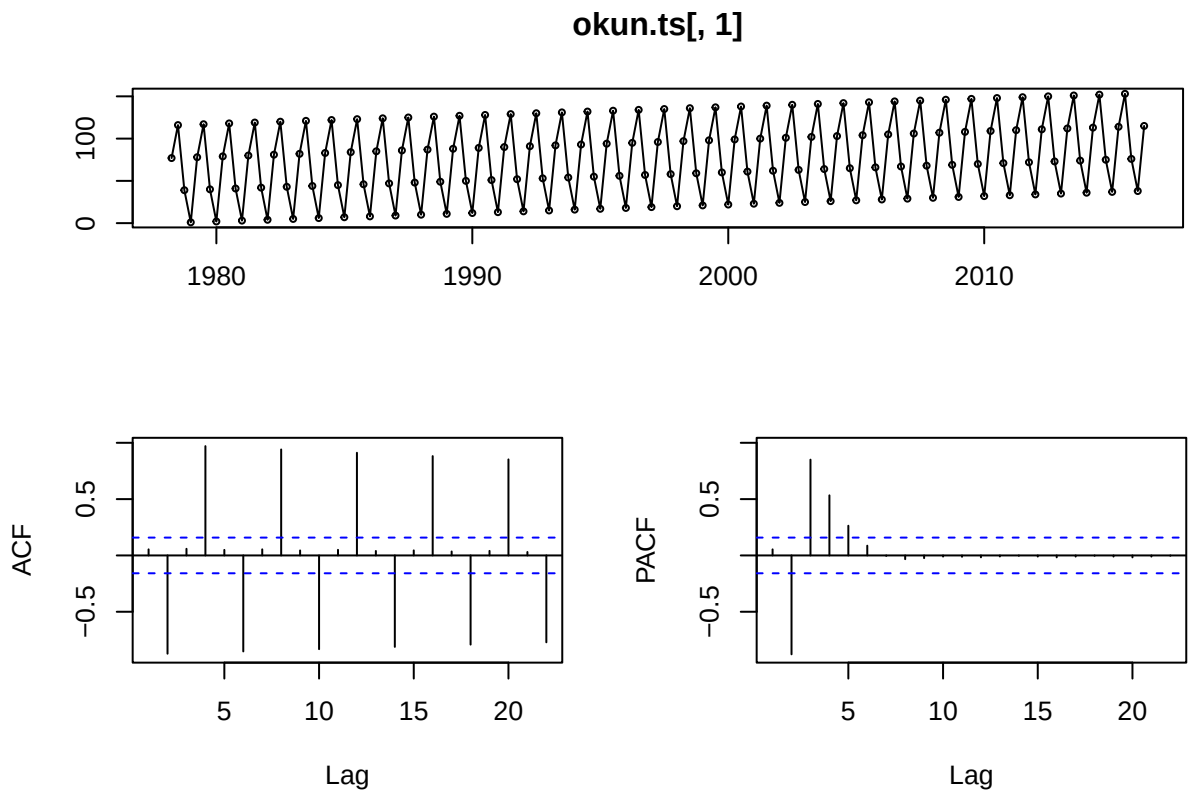
When you first obtain a time series, you want to convert it into a time series which uses the `tseries` package. You input: `converted <- ts(data, start = c(year, period), end = c(year, period), frequency = )`

DYNLM regression

In the `dynlm` package, we can difference using `d` and find lags using `L`. Say we'd like to create a model called regmod with the y variable being differenced and two lags of the x variable:

$$Dy_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + x_{t-2} + e_t$$

```
data("okun5_aus", package = "POE5Rdata")
okun.ts <- ts(okun5_aus, start=c(1978,2), end=c(2016,2),frequency=4)
okun.mod <- dynlm(d(u) ~ L(g,0:4), okun.ts)
tsdisplay(okun.ts[,1])
```




```
#or individually, using acf and pacf
```

Then you may carry out your analysis. How can we see what lagging or differencing a variable does? Take the growth column and lag it. View the head since it is a large dataset. What do the time series plots look like?

Stationarity Tests

Although it is helpful, it is not rigorous to simply observe the plots to determine stationarity. There must be some formal test: unit root test. The most common is the **Augmented Dickey-Fuller Test (ADF)**

a) $p-value > 0.05$: Fail to reject the null hypothesis (H_0), the data has a unit root and is non-stationary.
b) $p-value \leq 0.05$: Reject the null hypothesis (H_0), the data does not have a unit root and is stationary.

This can be done explicitly using the `adf.test` in `tseries`, or within the forescast package, you may consult `ndiffs` which does it for you. How many times should growth rate be differenced to be stationary in our okun5_aus datset? What about unemployment? Is there seasonality present? You can visually inspect using the time series plot or test using `nsdiffs`.

Classification of your Model

After conducting the above steps, you'll need to assign a lag order to your model. To double check if you have done so correctly, you may employ `auto.arima()` or `auto.sarima()`.

Testing for Autocorrelation

As previously mentioned, we may use the correlogram of the residuals, DW (if order 1) or BG test for residual auto correlation. Test the model given for autocorrelation by using the acf, and BG test of the appropriate order.