Agenda

1. Time Series
2. Moving Averages
3. Autoregressive models
4. ARIMA models
5. Putting it all together

Time Series A Time series is a set of variables measured over time. Some examples are census data taken at equally spaced intervals with a consistency in the method of measurement and the variables measured, meteorological measurements, utility supply and demand measurements.

Before we proceed with modeling any time series data we need to understand the mechanism generating the time series. These include describing any sources of variations, seasonality, trends, cycles inherent in a time series.

Trend Trend is a long term pattern in the series that is either decreasing or increasing. It can have a "zig-zag" pattern for shorter time periods within the series.

Seasonality Seasonality is a recurring pattern for each period e.g. annual seasonal patterns have a cycle that is 12 periods long, with each month being a period. If the periods are quarters then the annual seasonal pattern is 4 periods long. It is desirable to get an estimate of the seasonal index to assess seasonality in the dataset. A seasonal index measures the extent of seasonal influence for a particular segment of the year i.e. it compares the expected value of each period to the overall or grand mean. In particular, we are interested in estimating how much each value for a period is above or below the grand mean.

Note: Most time series data from US government websites is deseasonalized i.e. trends to seasonality factors have been removed.

Cycle These are patterns in the data which are not due to seasonality.

Decomposition analysis to separate out the Seasonal (S), Trend (T) and Cycle (C) components along with Irregularity (I) and remove their influence from the data (X) before analysis.

$$X_t = S_t T_t C t I$$

Moving Averages are smoothing techniques that can help discern the various components above in the data.

Moving Averages In classical time series, Trend-Cycles are removed using "moving averages". A simple moving average of order 'm" is defined as:

$$\widehat{T_t} = \frac{1}{m} \sum_{j=-k}^{k} X_{t+j}$$

Other moving averages methods include weighted moving averages, and other complex smoothing techniques.

Autoregressive models In an autoregressive model we use multivariate regression to model the variable using a linear combination of the variable values over past time periods. The model for order p is:

Consider the addition of an *interaction* term:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \epsilon_t$$

See Time series lab for example of AR(1) and AR(2) models and conditions.

Autoregressive moving average (ARIMA) Models ARIMA models are widely used to model time series data.

Most forecasting methods in time series also require certain stationarity conditions. Here are some commonly desirable conditions:

First Order Stationary: If the expected value of variable X(t) is the same for all t then it is said to be First Order Stationary. This is typically achieved by detrending the data using methods such as differencing.

Second Order Stationary: If X(t) is first order stationary and the covariance between X(t) and X(s) is a function of the length (t-s) then is is Second Order Stationary. In econometrics this is commonly achieved by taking a transformation e.g. taking square root of the variable.

Taking differences of two consecutive observations is one approach to making the time series stationary. Seasonal differencing is used when differences are taken over consecutive periods. More on differencing in R is illustrated in the lab handout.

Maximum likelihood estimation is the method of choice (you used this in logistic regression) as it is similar to the least squares regression estimate in that minimizes the sum of square residuals. Similar to logistic regression, here we find the parameter estimates that maximize the log likelihood. R uses AIC to find the "best" model. Corrected AIC along with details on estimation and order selection can be found in Section 8.6 of Forecasting: principles and practice

Putting it all together Lets consider the material we have covered throughout this course. We have tried to focus mostly on applied ideas, some of which are computational (more useful in practice, hard to test) and some which are slightly more theoretical (needed so you understand what you can apply, easier to test).

Lets try to categorize some of the tasks we have learned how to perform. Where in the CFAU process do these fit?

Interpreting multiple regression coefficients, nested F-tests, added variable plots, stepwise regression, calculating Cook's distance, calculating Fisher's LSD, calculating Tukey's HSD, the Bonferroni correction, randomization tests, bootstrap methods, fitting ANOVA, rewriting ANOVA models in different ways, checking ANOVA p-values, fitting a logistic regression model, translating from log-odds to odds, translating from odds to probability, calculating odds ratios, looking at p-values on logistic regression, interpreting logistic regression coefficients, considering forms of models.

Choose

Fit

Assess

Use

Because interpreting coefficients is so important for both your project and your test, lets practice writing out the "script" for a variety of cases.

1. Multiple regression: categorical variable

2. Multiple regression: numeric variable

3. Multiple regression: interaction term

4. ANOVA α_i

5. ANOVA μ_i

6. Logistic regression: categorical variable

7. Logistic regression: numeric variable

8. Logistic regression: interaction term