SARIMAX

Seasonal ARIMA with Exogenous Variables

Financial Time Series Final Project
May 11, 2022

Introduction

- We've thought about time series as models that use past values and errors of itself to forecast ahead
- Past values being autoregressive
- Past errors being moving average process
- Ex: Stock prediction
- Many predicting factors (labor costs, market demand etc.)
- Assumption that exogenous variables that affect the stock price will be close in value when close in temporal proximity allows for exogenous variables to not have to be directly included
- Possible that all/some of the exogenous variables are good predictors and can be included along with AR, MA components

SARIMAX

- Seasonal ARIMA with exogenous variables
- Since exogenous variables are indirectly manifested through past values, SARIMA would still work in place of SARIMAX
- The idea is to achieve better predictive power with the addition of exogenous variables
- Hard to find data that outright does not work for SARIMA but works for SARIMAX

SARIMAX - Formula

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

SARIMAX - Formula

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

- SARIMA component
- n exogenous variables at time t
- Exogenous variables not autoregressed or do any special operation
- Linear combination of variables and coefficients at time t that affect value of time series at point t

SARIMAX - Interpreting Coefficients

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + z_t,$$

$$y_t = \beta x_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + z_t$$

- ARMA(p,q) vs ARMAX(p,q)
- Cannot be thought of as combining regression and time series because beta coefficient is interpreted differently
- Classic regression = beta is effect on Yt when Xt increased by one
- ARMAX = beta is interpreted conditionally on past values of Yt
- Not as intuitive as regression

SARIMAX - Forecasting Warning/Caveats

- ARIMA/SARIMA allows us to forecast many timesteps into the future
- What about SARIMAX?
- SARIMAX includes exogenous variables at present time t to forecast t+1
- What about 2 or 3,4… timesteps into the future? (t+2, t+3…)
- Exogenous variables only known at time t, to forecast t+2 need to know exogenous values at t+1
- Exogenous variables need to be forecasted in order for time series to be forecasted
- May also be a time series itself, or could follow some other model (linear, polynomial, Poisson etc.)
- Every forecast has possibility for error. Many forecasts (several exog. variables, predicting many timesteps into future) compounds error and can magnify the error in time series forecast

SARIMAX - Forecasting Limitations

- Use SARIMAX when predicting only 1 timestep in the future is all that is needed
- Choose exogenous variables that have low prediction error
- Less chance errors magnify when exog. variables can be accurately predicted
- Ex. SARIMAX model for forecasting daily temperature highs
- Time of sunrise and sunset as exog. variables

Sources

https://robjhyndman.com/hyndsight/arimax/

https://phosgene89.github.io/sarima.html

Time Series Forecasting in Python - Marco Peixeiro

Manning Publications 2021

Python Demo

- Walmart sales tracking 50 items sold in 10 different stores over 5 years
- Focused on sales of one item at one store location