Time Series Analysis & Forecasting

Class 6

Arnab Bose, Ph.D.

MSc Analytics

University of Chicago

Cross-Validation

- Model validation technique
- Divide a dataset into a training set and a validation set
- K-fold cross validation => original data is divided into K equal size subsamples. Of the K, 1 subsample is retained as the validation set. Remaining K-1 are used as the training data
- The above is repeated K times with each K subsample used exactly once as the validation data
- The results from the validations will be combined (eg. average) to produce a single model estimation

Cross-Validation

- Take a rolling window for train, set forecast with h steps ahead
- Calculate accuracy
- Roll window forward by h steps and repeat
- https://robjhyndman.com/hyndsight/tscv/
- Challlenges and nested CV: https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

R code – TS CV

- *data(gas)*
- *auto.arima(gas)*
- *far2 <- function(x, h){forecast(Arima(x, order=c(2,1,1), seasonal = c(0,1,1)), h=h)}*
- *e <- tsCV(gas, far2, h=3)*
- *eDf <- as.data.frame(e)*
- *matplot(eDf, type="l")*

Regression with ARMA errors

- Regression assumption that errors are not autocorrelated is violated

$$y_t = \beta_0 + \beta_1 x_t + u_i$$

where u_i is autocorrelated

- So u_i can be modeled as

$$\phi(B)u_i = \theta(B)e_t$$

where e_t is white noise

- ARMA can be considered as a special type of regression model => predictors are lags of the dependent variable and/or lags of the forecast error

R code – Regression with ARMA errors

- *x <- 1:100*
- *e <- arima.sim(model=list(ar=0.3, ma=0.9), n=100)*
- *y <- 1 + 2*x + e*
- *fit1 <- lm (y~x)*
- *summary(fit1)*
- *Plot(fit1)*
- *acf(fit1$residuals, lag=100)*
- *par(mfrow=c(1,2))*
- *qqnorm(fit1$residuals)*
- *qqline(fit1$residuals)*
- *fit2 <- auto.arima(y, xreg=x)*
- *summary(fit2)*
- *qqnorm(fit2$residuals)*
- *qqline(fit2$residuals)*
- *acf(fit2$residuals)*

Long Memory Fractional ARIMA – ARFIMA

- Compromise between short term memory ARMA models and the fully integrated ARIMA models

- Consider the ARIMA model

$$\Phi(B)(1-B)^d y_t = \Theta(B)e_t$$

- Here |d| < 0.5
  - TS $\{y_t\}$ is stationary if d < 0.5
  - TS $\{y_t\}$ is invertible if d > -0.5

- ACF goes to infinity for 0 < d < 0.5 => long term memory

- ACF decays hyperbolically to 0 for -0.5 < d < 0 => intermediate memory

R code – ARFIMA

- *library("arfima", lib.loc="~/R/win-library/3.3")*
- *data("SeriesJ")*
- *acf(SeriesJ$YJ, lag=40)*
- *library("forecast", lib.loc="~/R/win-library/3.3")*

- *d <- fracdiff::fracdiff(SeriesJ$YJ) #get the fractional d*
- *st <- fracdiff::diffseries(SeriesJYJ,dd) # do the fractional difference*
- *acf(st, lag=40)*
- *m1 <- auto.arima(st) # now the TS is stationary, run ARIMA*
- *AIC(m1)*

- *m2 <- forecast::arfima(SeriesJ$YJ) # does the above (fractional difference + ARIMA) in 1 step*
- *AIC(m2) # you can compare the models and choose the one with lower AIC*

Model Uncertainty

- Sequential Stability
  - Take different sample sizes and calculate the % of times same model selected

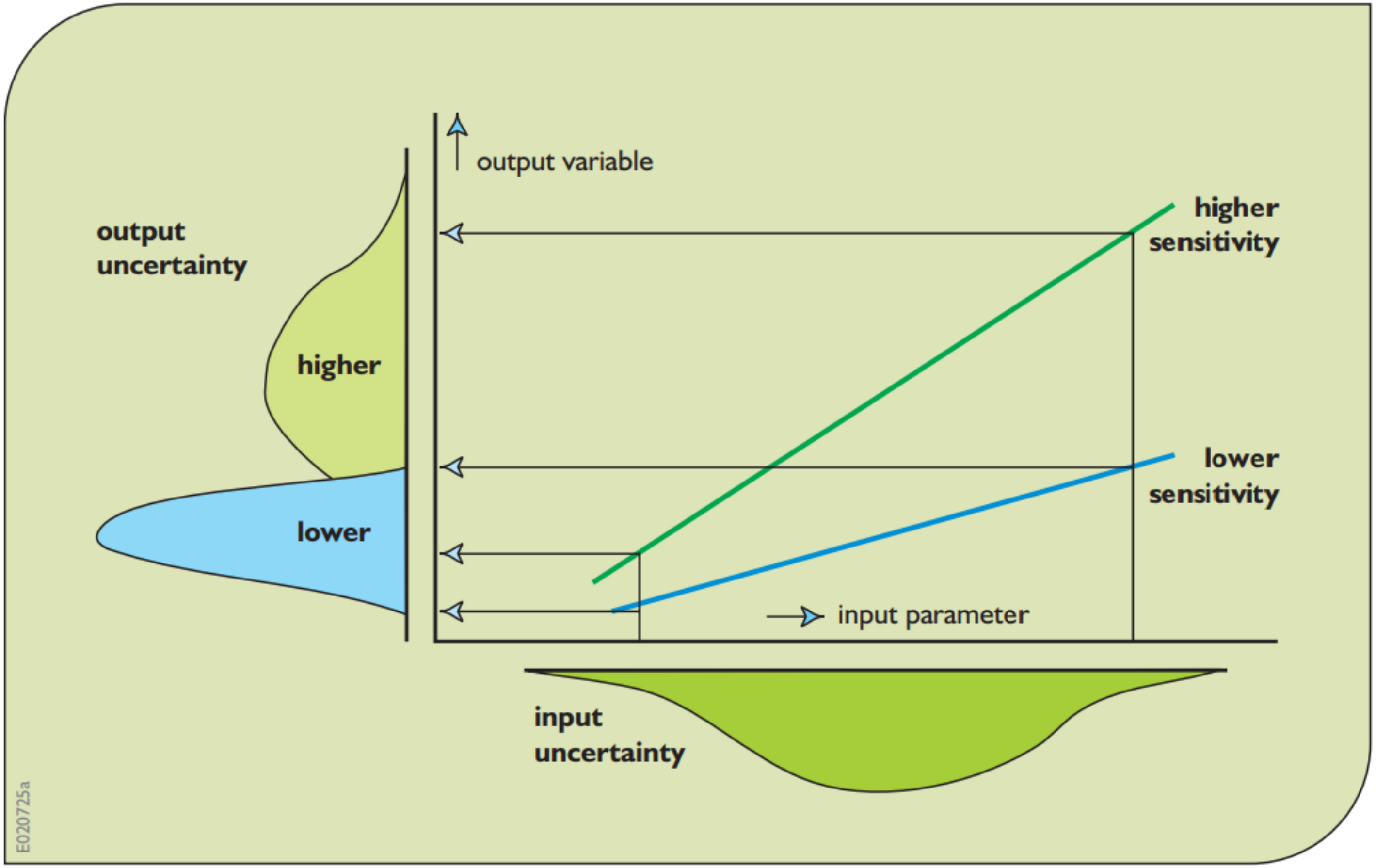
- Perturbation Stability
  - Change dataset with new additional samples and selected model should remain the same

- NOTE – Model selection method does not guarantee any forecasting performance

Sensitivity Analysis

- Similar to risk, sensitivity assesses how the uncertainty affects area of interest for a particular use case.
- Measures the sensitivity of a model output to changes in model input values.
- Input/output scatterplots are a simple method for sensitivity analysis.
- To calculate model output sensitivity analysis to different input parameters, use Monte Carlo simulations.
- To focus on specific regions, use Monte Carlo Filtering.

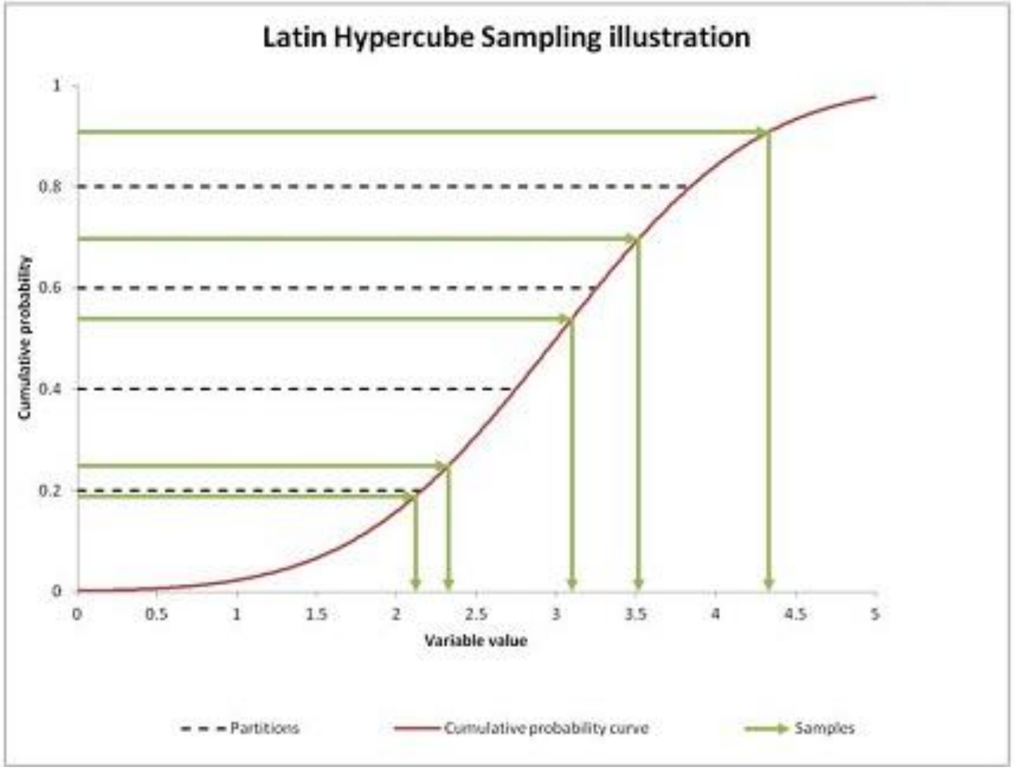
Schematic diagram for input param uncertainty and sensitivity



LAL, W. 1995. Sensitivity and uncertainty analysis of a regional model for the natural system of South Florida. West Palm Beach, Fla., South Florida Water Management District. Draft report, November.

https://ecommons.cornell.edu/bitstream/handle/1813/2804/09_chapter09.pdf;sequence=12

Latin Hypercube Sampling



http://liprof.com/blog/the-pros-and-cons-of-latin-hypercube-sampling

Textbook Chapters

- Materials covered available in book:
  - FPP: Chapter 9, MKJ: Chapter 4, TSA: Chapter 9