# Time Series Analysis & Forecasting

# Class 2

Arnab Bose, Ph.D.

MSc Analytics

University of Chicago

# Non-uniform to Uniform TS

Approaches:

1. Forward fill
2. Backward fill
3. Interpolate (linear, spline, ..)

# Trend Stationary

- Deterministic Trend

$$y_t = \beta t + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2)$$

$$\mu = E[y_t] = \beta t$$

$$Var(y_t) = E[(y_t - \mu)^2] = E[\varepsilon_t^2] = \sigma^2$$

# Difference Stationary

- Stochastic Trend

$$y_t = \beta + y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2)$$

$$\mu = E[y_t] = \beta t$$

$$Var(y_t) = E[(y_t - \mu)^2] = E[t\varepsilon_t^2] = t\sigma^2$$

# R code – Trend & Difference Stationary

```
# Trend
library(tseries)
wn <- rnorm(500)
wnt <- wn + 0.3*(1:length(wn))
ts.plot(wnt)
adf.test(wnt)# detects trend stationarity
kpss.test(wnt)
kpss.test(wnt, null="Trend")# need to explicitly mention trend stationarity
# Difference - Random Walk with Drift
wnd <- wn + 0.9
rwd <- cumsum(wnd)
ts.plot(rwd)
adf.test(rwd)
kpss.test(rwd)
adf.test(diff(rwd))
kpss.test(diff(rwd))
```

# Holt-Winters (Additive)

HW is a smoothing algorithm for a TS $\{y_t\}$ that exhibits linear trend and seasonality changing slowly with time. It is a smoothing algorithm with constants $0 \le \alpha, \beta, \gamma \le 1$ and periodicity $L$. Prefer not to use with stationary data as it is sub-optimal from accuracy perspective.

- Additive model to be used when the seasonal variation is additive in nature – toy sales increase by $1 million every Dec.

Level

$$l_t = \alpha(y_t - s_{t-L}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

Trend

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

Seasonality

$$s_t = \gamma(y_t - l_{t-1}) + (1-\gamma)s_{t-L}$$

Forecast

$$\hat{y}_{t+\tau}(t) = l_t + \tau b_t + s_{t+\tau-L}$$

# Holt-Winters (Multiplicative)

- Multiplicative model to be used when the seasonal variation is multiplicative in nature – toy sales increase by 42% every Dec.

Level

$$l_t = \alpha(y_t/s_{t-L}) + (1-\alpha)(l_{t-1} + b_{t-1})$$

Trend

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$

Seasonality

$$s_t = \gamma(y_t/l_{t-1}) + (1-\gamma)s_{t-L}$$

Forecast

$$\hat{y}_{t+\tau}(t) = (l_t + \tau b_t)s_{t+\tau-L}$$

# Holt-Winters with Damped Trend

Introduces a damping parameter $0 \leq \phi \leq 1$ that changes the level and trend component and therefore, the forecast.

<u>Additive</u>

Level

$$l_t = \alpha(y_t - s_{t-L}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$$

Trend

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$$

Forecast

$$\hat{y}_{t+\tau}(t) = l_t + (\phi + \phi^2 + \phi^3 + \cdots + \phi^\tau) b_t + s_{t+\tau-L}$$

<u>Multiplicative</u>

Level

$$l_t = \alpha(y_t / s_{t-L}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$$

Trend

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)\phi b_{t-1}$$

Forecast

$$\hat{y}_{t+\tau}(t) = (l_t + (\phi + \phi^2 + \phi^3 + \cdots + \phi^\tau) b_t) s_{t+\tau-L}$$

# R code

```
require(graphics)
## Seasonal Holt-Winters
m <- HoltWinters(co2)
plot(m)
plot(fitted(m))
m <- HoltWinters(AirPassengers, seasonal = "mult")
plot(m)

## Non-Seasonal Holt-Winters
x <- uspop + rnorm(uspop, sd = 5)
m <- HoltWinters(x, gamma = FALSE)
plot(m)

## Exponential Smoothing
m2 <- HoltWinters(x, gamma = FALSE, beta = FALSE)
plot(m2)
```

# Regression

- Introduced by Sir Francis Galton in “Family Likeness in Stature” 1886.
- Estimate or forecast the average value of one variable (dependent) on the basis of fixed values of other variables (independent)
- Dependent ≡ Explained ≡ Predictand ≡ Regressand ≡ Response ≡ Endogeneous ≡ Outcome ≡ Controlled
- Explanatory ≡ Independent ≡ Predictor ≡ Regressor ≡ Stimulus ≡ Exogeneous ≡ Covariate ≡ Control

# Regression – Assumptions

Mathematical representation $y_t = \widehat{\beta_0} + \widehat{\beta_1} x_t + u_t$

1. Linear in the parameters
2. X values are fixed in repeated sampling
3. Zero mean value of the disturbance $u_t$
4. Homoscedasticity or equal variance of $u_t$
5. No autocorrelation in $u_t$
6. Zero covariance between $u_t$ and $x_t$
7. Number of observations must be greater than number of parameters to be estimated
8. X values in a given sample must vary
9. No specification bias in the model
10. No perfect linear relationships among the explanatory variables

# Regression – what about error normality assumption ?

- Probability distribution of the parameters depend on the assumption of the probability distribution of $u_t$
- NOTE – OLS makes no assumption about the probability of $u_t$
- Advantage of normal distribution of $u_t$, i.e. normal distribution of the parameters allow the use of t, F and $\chi^2$ distributions
- If normal, then OLS and Maximum Likelihood (ML) estimators are identical.

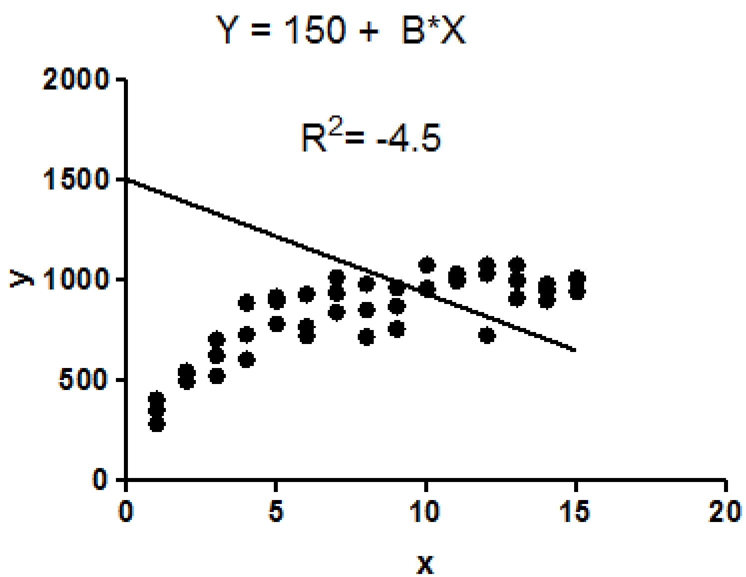
# Multiple Regression

Mathematical representation $y_t = \widehat{\beta_0} + \widehat{\beta_1} x_{1t} + \widehat{\beta_2} x_{2t} + u_t$

- Pearson correlation among the variables: $r_{yx1}, r_{yx2}, r_{x1x2}$
- These correlation coefficients are known as zero-order correlations because they do not control for interconnection amongst variables.
- Multiple coefficient of determination is $R^2$ that represents how much of the variability in $y$ can be explained by the predictor variables.
- Highest value of $R^2$ is given by $R^2 = r_{yx1}^2 + r_{yx2}^2$

# Negative $R^2$

$$R^2 = 1 - \frac{variance(residuals)}{variance(outcomeAboutMean)}$$




https://stats.stackexchange.com/questions/12900/when-is-r-squared-negative

# Multicollinearity

- If present,
  - OLS estimators will have large variances and covariances
  - the confidence intervals of the estimators tend to be wider – readily accept the null hypothesis
  - Estimators will be sensitive to small changes in data – high standard error
  - Low t values for the estimators
- To detect, use the Variation Inflation Factor denoted by $VIF = \frac{1}{1 - r_{x1x2}^2}$
- VIF indicates how the variance of the estimator is inflated by multicollinearity

# R code (continued)

```
require(car)
fit <- lm(prestige ~ income + education, data=Duncan)
fit
vif(fit)

x <- c(1,2,3,4,5)
y <- c(11,20,25,24,29)
mod <- lm (y ~ x)
mod0 <- lm(y ~ x + 0)
summary(mod)
summary(mod0)
plot(x, y, ylim=c(0, 30))
abline(mod, col="blue")
abline(mod0, col="red")
```

# Data Transformations

- Box-Cox transformation $w_t$ of TS $y_t$

$$w_t = \begin{cases} \log(y_t) \;\; if\ \lambda = 0 \\ \dfrac{y_t^{\lambda} - 1}{\lambda} \;\; otherwise \end{cases}$$

- Use a transformation to
  - Decouple mean and variance to remove variance dependence on mean
  - Model is simple (additive)
  - Residuals are more or less normally distributed with zero mean and constant variance
- Note that the transformation is not a guarantee for normality – it actually does not check for normality but for the smallest standard deviation.
- Assumption is that the transformed data has the highest likelihood (but not a guarantee) t0 be normally distributed when the standard deviation is the lowest.
- Also this transformations only works if all the data is positive and > 0.

# R code

```
library(forecast)
lambda <- BoxCox.lambda(lynx)

lynx.fit.arima.boxcox <- auto.arima(lynx, lambda=lambda)
plot(forecast(lynx.fit.arima.boxcox, h=20))
lambda_separate <- BoxCox.lambda(lynx)
lynx.transformed <- BoxCox(lynx, lambda_separate)
lynx.transformed.fit.arima <- auto.arima(lynx.transformed)
plot(forecast(lynx.transformed.fit.arima, h=20))
```

# Wold Decomposition

A (weak) stationary TS can be written as the sum of 2 TS, one deterministic and one stochastic

$$Y_t = \sum_{i=0}^{\infty} \psi_i \, w_{t-i} + \eta_t$$

Where

$\eta_t$ is deterministic process

$\psi_0 = 1 \text{ and } \sum_{i=0}^{\infty} {\psi_i}^2 < \infty$ (process is stable)

$w_t$ is white noise

$Cov(w_t, \eta_t) = 0 \text{ for all } t, s$

# Simple Operators

- Backward shift operator $B\, z_t = z_{t-1}$
- Forward shift operator $F\, z_t = z_{t+1}$
- Backward difference operator $\nabla z_t = z_t - z_{t-1}$

# Linear Process

- TS $\{y_t\}$ is a weighted linear representation of independent shocks $\{e_t\}$ that represent white noise
- Assume $\{e_t\}$ to be normally distributed with mean 0 and variance $\sigma_e^2$
- Mathematically the TS can be represented as

$$y_t = \mu_t + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots$$

where

$\mu_t$ - determines the "level" of the process

$\psi_t$ - constant coefficients that need to be summable for stationarity

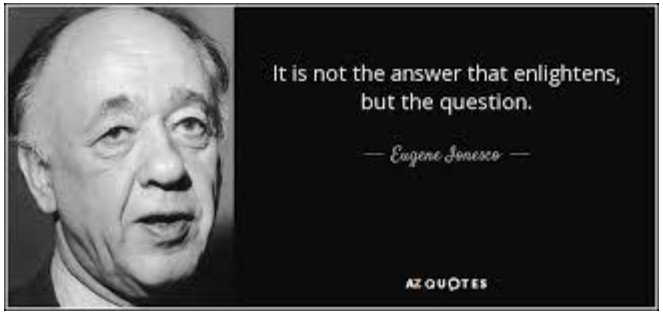
# Textbook Chapters

Materials covered available in book chapters:

FPP: 7 – 9, 10

PTS: 2, 6

# Thank You


It is not the answer that enlightens,
but the question.
— Eugene Ionesco —
AZ QUOTES