Formula Sheet

Jeff Nguyen

April 2, 2021

Gauss-Markov Theorem

6th Assumptions: Stationarity

A variable if overtime, the mean μ_X, variance σ_X and covariance $Cov(Y_t, Y_{t-5})$ stays the same.
- If the mean changes: it's a "mean non-stationary" variable.
- If the variance changes: it's a "variance non-stationary" variable.

Review the regression equation:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \tag{1}$$

Y	X	ϵ	Regression OK?	Note
Stationary	Stationary	Stationary	Yes	
Non-Stationary	Non-Stationary	Stationary	Yes	Cointegration, needs more data
Non-Stationary	Non-Stationary	Non-Stationary	No	Spurious Regression

Time Trend

Every time series variable can be decomposed to the following trend:
- Time
- Seasonal
- Cyclical
- Autocorrelation
- Randomness

Note: Recession is usually an economic variable.

How to measure time trend?

Step 1:

Run Regression:

$$
\begin{gathered}
Y = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta T^3 + \epsilon \\
\text{Note: Nothing beyond } T^3
\end{gathered}
\tag{2}
$$

If T^3 is insignificant, drop it and so on. The trend estimated equation is:

$$
\begin{aligned}
\hat{Y} &= \hat{\beta_0} + \hat{\beta_1} T + \hat{\beta_2} T^2 + \hat{\beta_3} T^3 \\
Y &= \hat{Y} + \epsilon
\end{aligned}
\tag{3}
$$

$\hat{\epsilon}$ represents seasonality, cyclicality, autocorrelation, and randomness.

Step 2:

Use $\hat{\epsilon_t}$ to estimate autocorrelation:

$$
\hat{\epsilon_t} = \alpha_0 + \alpha_1 \hat{\epsilon_{t-1}} \tag{4}
$$

What to do when you have spurious regression, i.e. ϵ is non-stationary?

In this case, ϵ shows as not scattered on a scatterplot. We have to de-trend the variable.

Two type of trend:
- Deterministic: when variable explicitly driven by trend (saving account balance).
- Stochastic: random variable (S&P 500 price movement).

De-trend Deterministic Trend Variable:

- Run a regression on time: $Y = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta T^3 + \epsilon$.
- The estimated $\hat{Y}$ is

$$
\hat{Y} = \hat{\beta_0} + \hat{\beta_1} T + \hat{\beta_2} T^2 + \hat{\beta_3} T^3 \tag{5}
$$

- The equation is then $Y = \hat{Y} + \hat{\epsilon}$.
- Take the error of regression $\hat{\epsilon}$ as “de-trended” Y.

De-trend Stochastic Trend Variable:

- Differencing Y:

$$
\Delta Y = Y_t - Y_{t-1} \tag{6}
$$

For example, stock price model:

$$
\begin{aligned}
Y_t =& Y_{t-1} + \epsilon_t \\
& \text{via inspection:} \\
Y_t =& Y_0 + \sum_{i=0}^{t} \epsilon_i \\
& \text{thus:} \\
E[Y_t] =& Y_0 \quad \text{The model is mean-stationary.} \\
Var[Y_t] =& t\sigma_\epsilon^2 \quad \text{Variance non-stationary}
\end{aligned} \tag{7}
$$

Note that the random-walk hypothesis is variance-non-stationary.

Removing "wrong" Nonstationarity:

- If the variable is Trend Stationary (TS) and was treated as Difference Stationary (DS), the estimated $\hat{\beta}$ will be unbiased, inefficient and consistent. Use more data to fix the problem.
- If the variable is DS and was treated as TS, $\hat{\beta}$ will be biased, inefficient and inconsistent.

Smoothing Technique

Simple Moving Average (SMA)

Moving Average is not a point forecast, it's a trend forecast.

How to calculate

Y	$MA(3)$
3	
9	$\frac{3+9+6}{3} = 6$
6	$\frac{9+6+7}{3} = 7.33$
7	$\frac{6+7+5}{3} = 6$
5	$\frac{7+5+2}{3} = 5.33$
2	

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(pracma)

# Code to calculate simple MA
```

```r
yArray <- c(3, 9, 6, 7, 5, 4, 2)
n = 3

smaArray <- movavg(yArray, n=n, type="s")
startloc <- ceiling((n+1)/2)
startloc
```

```
## [1] 2
```

```r
smaArray[startloc+1:length(smaArray)]
```

```
## [1] 6.000000 7.333333 6.000000 5.333333 3.666667       NA       NA
```

Centered Moving Average (CMA)

CMA is used to de-seasonalize data.

Y	$MA(4)$
3	
1	
0	$\frac{3+1+0+4}{4} = 2$
4	$\frac{1+0+4+2}{4} = 1.75$
2	$\frac{0+4+2+1}{4} = 1.75$
1	

```r
library(forecast)
library(pracma)

# Code to calculate simple MA
yArray <- c(3, 1, 0, 4, 2, 1)
n = 4

smaArray <- movavg(yArray, n=n, type="s")
loc1 <- ceiling((n+1)/2)
loc2 <- ceiling((n+1)/2)
startloc <- (loc1+loc2)/2

startloc
```

```
## [1] 3
```

```r
smaArray[startloc+1:length(smaArray)]
```

```
## [1] 2.00 1.75 1.75   NA   NA   NA
```

Weighted Moving Average (WMA)

$$\begin{aligned} Y_{t+1} &= .5Y_t + .3Y_{t-1} + .2Y_{t-2} \\ \sum w &= .5 + .3 + .2 = 1 \end{aligned} \tag{8}$$

Note: assign more weights to most recent data.

```
yArray <- c(3, 9, 6, 7, 5, 4, 2)
n = 3

movavg(yArray, n=n, type="w")
```

```
## [1] 3.000000 7.000000 6.500000 7.000000 5.833333 4.833333 3.166667
```

Problem of Moving Average

- Losing of observations (for example: beginning and end of data) and degree of freedom.
- Weights are equal weights when in reality, more weights should be given to most recent data.
- It's not a point forecast.

Remarks: Exponential Smoothing solve these problems based on the idea that the forecast of future time period is based on:

$$
\begin{aligned}
Y_{t+1} =& Y_t^F + \alpha(Y_t - Y_t^F) \\
& \text{where } 0 < \alpha < 1 \text{ is the damping effect} \\
& \text{therefore:} \\
Y_{t+1}^F =& \alpha Y_t + (1-\alpha) Y_t^F \\
=& \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots
\end{aligned}
\tag{9}
$$

Reviewing Forecasting:

ME = Mean Error (Average of errors).
RMSE = Root Mean Square Error (SD of residuals, i.e. how far data is from regression line).
MAE = Mean Absolute Error (Average of distance from regression line).
MAPE = Mean Absolute Percentage Error (measure prediction accuracy).
MASE = Mean Absolute Scaled Error (prediction accuracy).

When decompose time series, use additive for log variables and multiplicative for others.

Simple Exponential Smoothing

Use if data is flat.

Pick the damping effect that yields the smallest ACF1 metric.

Exercise: Given damping effect α, do one-period Exponential Smoothing forecast

Use the formula $Y_{t+1} = Y_t^F + \alpha(Y_t - Y_t^F)$. Note that it's not possible to do more than one-period ahead forecast using ES.

Confidence Interval:

$$
Y^F - 1.96\epsilon_F \leqslant Y_{actual} \leqslant Y^F + 1.96\epsilon_F \tag{10}
$$

Holt Trend Model

Use if data has time trend use Holt Exponential Smoothing. If data has seasonality, use Holt-Winter Exponential Smoothing.

$$\begin{aligned}
\text{Forecast Equation:}& \\
Y_{t+h}^F =& l_t + hb_t \\
\text{Level Equation:}& \\
l_t =& \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
\text{where}\,\alpha\,\text{is the smoothing coefficient for level}& \\
\text{Trend Equation:}& \\
b_t =& \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\
\text{where}\,\beta\,\text{is smoothing coefficient for trend}&
\end{aligned} \tag{11}$$

Random Walk

Random Walk (Naive Forecasting)

$$\begin{aligned}
Y_t =& Y_0 + \sum_{i=1}^{n} \epsilon_i \\
\mu =& E[Y_t] = Y_0 \quad \Leftarrow \quad \text{stationary} \\
\sigma^2 =& Var[Y_t] = t\sigma_\epsilon^2 \quad \Leftarrow \quad \text{non-stationary}
\end{aligned} \tag{12}$$

Random Walk with Drift

$$\begin{aligned}
Y_t =& Y_0 + t\alpha_0 + \sum_{i=1}^{T} \epsilon_i \\
\mu =& E[Y_t] = Y_0 + t\alpha_0 \quad \Leftarrow \quad \text{stationary} \\
\sigma^2 =& Var[Y_t] = t\sigma_\epsilon^2 \quad \Leftarrow \quad \text{non-stationary} \\
& \text{and} \\
D(Y_t) =& \epsilon_t \\
E[DY_t] =& 0 \\
Var[DY_t] =& \sigma_\epsilon^2
\end{aligned} \tag{13}$$

Random Walk with Drift and Noise

$$\begin{aligned}
Y_t =& \alpha_0 + Y_{t-1} + \epsilon_t + \eta_t \\
E[Y_t] =& Y_0 + \alpha_0 t \\
Var[Y_t] =& \sigma_\epsilon^2 + \sigma_\eta^2
\end{aligned} \tag{14}$$

1st Order Linear Homogenous DE

$$\begin{aligned}
Y_t =& \alpha Y_{t-1} \quad \text{given}\, Y_0 \\
\text{Solution:} \quad Y_t =& \alpha^t Y_0
\end{aligned} \tag{15}$$

If $0 < \alpha < 1$**:** Y is a non-oscillating convergent model (stationary).
If $\alpha > 1$**:** Y is a non-oscillating divergent model (non-stationary).
If $-1 < \alpha < 0$**:** Y is an oscillating convergent model.
If $\alpha < -1$**:** Y is an oscillating divergent model.

1st Order Linear Non-Homogenous DE

$$\begin{aligned} Y_t =& \alpha_0 + \alpha_1 Y_{t-1} \quad \text{given } Y_0 \\ \text{Solution:} \quad Y_t =& \alpha_0 \frac{1-\alpha_1^t}{1-\alpha_1} + \alpha_1^t Y_0 \end{aligned} \tag{16}$$

This result is used in Box-Jenkins methodology and ARIMA model.

ARIMA Model

Note that ARIMA has "short memory" because of its recursive nature $\Rightarrow$ each iteration erodes efficiency.

Examples:

$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \epsilon_t \Rightarrow$ **ARIMA(1,0,0) or AR(1)**

$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \epsilon_t \Rightarrow$ **ARIMA(1,1,0)**

$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \epsilon_t + \beta\epsilon_{t-1} \Rightarrow$ **ARIMA(1,0,1) or ARMA(1,1)**

$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} \Rightarrow$ **ARIMA(1,1,2)**

$DY_t = \alpha_0 + \alpha_1 Y_{t-1} + \epsilon_t + \beta\epsilon_{t-1} \Rightarrow$ **Mis-specified model as DY_{t-1} is missing.**

Stationarity & Invertibility

Stationary if $\sum_{i=1}^n |\alpha_i| < 1$ and vice versa. Note that if $|\alpha_i| < 1$ and $\sum_{i=1}^n |\alpha_i| > 1$, non-stationarity is a possibility and an (A)DF Test is appropriate.

Invertibile if $\sum_{i=1}^n |\beta_i| < 1$ and N/A if there is no β_i.

Quick Calculations

AR(1) Model:

```
y_0 <- 2
sigmaSq <- 1.21
alpha_0 <- 1.6
alpha_1 <- .75

f <- function(y_t) {alpha_0 + alpha_1*y_t}
y_1 <- f(y_0)
y_2 <- f(y_1)
y_3 <- f(y_2)
```

Forecast Variable	Forecast Value	95% Confidence Interval Equation	95% Confidence Interval
$Y_1^F = -.5Y_0$	3.1	$Y_{t+1}^F \pm 1.96\sqrt{\sigma^2}$	$P(0.944 \leqslant Y_{t+1} \leqslant 5.256) = .95$
$Y_2^F = -.5Y_1$	3.925	$Y_{t+2}^F \pm 1.96\sqrt{\sigma^2(1+\alpha^2)}$	$P(1.23 \leqslant Y_{t+2} \leqslant 6.62) = .95$
$Y_3^F = -.5Y_2$	4.54375	$Y_{t+3}^F \pm 1.96\sqrt{\sigma^2(1+\alpha_1^2+\alpha_1^4)}$	$P(1.5884518 \leqslant Y_{t+3} \leqslant 7.4990482) = .95$

```
yLR <- alpha_0 / (1 - alpha_1)
sigmaSq_yLR <- sigmaSq / (1 - alpha_1^2)
```

Forecast Long Run	Forecast Value	95% Confidence Interval Equation	95% Confidence Interval
$Y_{LR} = \frac{1.6}{0.75}$	6.4	$Y_{LR} \pm 1.96\sqrt{\frac{\sigma^2}{1-\alpha_1^2}}$	$P(3.1404344 \leqslant Y_{LR} \leqslant 9.6595656) = .95$

AR(2) Model:

```
alpha_0 <- 6
alpha_1 <- .7
alpha_2 <- .12
y_0 <- 5
y_1 <- 6
sigmaSq <- 1.21
f <- function(y_tMinus1, y_tMinus2) {alpha_0 + alpha_1*y_tMinus1 + alpha_2*y_tMinus2}
y_2 <- f(y_1, y_0)
y_3 <- f(y_2, y_1)
y_4 <- f(y_3, y_2)
```

Forecast Variable	Forecast Value	95% Confidence Interval Equation	95% Confidence Interval
$Y_2^F = 6 + 0.7Y_1 + 0.12Y_0$	10.8	$Y_{t+1}^F \pm 1.96\sqrt{\sigma^2 * (1+\alpha_1^2)}$	$P(8.1682666 \leqslant Y_{t+1} \leqslant 13.4317334) = .95$
$Y_3^F = 6 + 0.7Y_2 + 0.12Y_1$	14.28	$Y_{t+2}^F \pm 1.96\sqrt{\sigma^2(1+\alpha_1^2+\alpha_2^2+\alpha_1^4+2\alpha_1^2\alpha_2)}$	$P(11.3379486 \leqslant Y_{t+2} \leqslant 17.2220514) = .95$
$Y_4^F = 6 + 0.7Y_3 + 0.12Y_2$	17.292	$Y_{t+3}^F \pm 1.96\sqrt{\sigma^2[1+\alpha_1^2+(\alpha_1^2+\alpha_2)^2+(\alpha_1^3+2\alpha_2\alpha_1)^2]}$	$P(14.1504328 \leqslant Y_{t+3} \leqslant 20.4335672) = .95$

```
yLR <- alpha_0 / (1 - alpha_1 - alpha_2)
sigmaSq_yLR <- (1 - alpha_2)*sigmaSq / ((1 + alpha_2)*((1 - alpha_2)^2 - alpha_1^2))
```

Forecast Long Run	Forecast Value	95% Confidence Interval Equation	95% Confidence Interval
$Y_{LR} = \frac{6}{1-0.7-0.12}$	33.3333333	$Y_{LR} \pm 1.96\sqrt{\frac{(1-\alpha_2)*\sigma^2}{(1+\alpha_2)[(1-\alpha_2)^2-\alpha_1^2]}}$	$P(29.7497601 \leqslant Y_{LR} \leqslant 36.9169066) = .95$

Box-Jenkins' Methodology

Objective

Estimating the data generating process of a time-series variable as a function of time $Y_t = f(t)$.

Why ARIMA Model

- When structural modes are not useful: Daily stock price.
- Sometimes structural models are not good forecasting models.
- Sometimes data are not available for structural models.

Three Stages of ARIMA:

- **Identification:** Finding order of $ARIMA(p, d, q)$ using correlogram.
- **Estimation**
- **Diagnostics**

Two Rquirements of ARIMA:

- **Stationarity:** The variable should be tested for stationarity. If the variable is non-stationary, it should be converted to stationary by first differencing the variable.
- **Parsimony:** Selecting the model which is parsimonious (simple but have great explaining power) using AIC, SBC. Models which are not parsimonious may fit better to noise than signals. They will not forecast future well.

Measurement of Goodness in ARIMA:

AIC and SBC (not R^2) are measurement of goodness of fit for ARIMA estimates, with SBC emphasizing parsimony. A lower AIC or BIC value indicates good fit.

$$
\begin{aligned}
AIC =& \ln \sigma^2 + \frac{2k}{T} \\
SBC =& \ln \sigma^2 + \frac{k}{T} \ln T \\
& \text{where} \\
k =& p + q + 1 \\
T =& \text{Number of observations}
\end{aligned}
\tag{17}
$$

When run ARIMA model in R, to determine if coefficient is significant or not, divide coefficient by the S.E.. If result is large then it is significant.

Diagnostics:

ARIMA will have a good forecasting output if:
- The residual of the regression is white noise. To test for white noise error, use correlogram of the residual.
- Coefficients of ARIMA are stable. To test for the stability of the coefficients, split the data, run two ARIMA and do an F test.

F Test of Diagnostics:

- Split data to two parts, fit ARIMA to each part and to the whole sample.

$$
\begin{aligned}
F =& \frac{\frac{SSR_T - SSR_1 - SSR_2}{k}}{\frac{SSR_1 + SSR_2}{T - 2k}} \\
H_0 :& \text{Coefficients are the same.} \\
H_a :& \text{Coefficients are not the same.} \\
& \text{where} \\
SSR_T =& \text{Total Sum Squared Residual} \\
SSR_1 =& \text{SSR for first part of the sample} \\
SSR_2 =& \text{SSR for second part of the sample} \\
K =& p + q + 1 \\
T =& \text{Sample size}
\end{aligned}
\tag{18}
$$

Properties of a Good ARIMA Model:

- Parsimonious.
- Has coefficients that are stationary and are invertible.
- Fits data well.
- Has residuals that are white noise.
- Has coefficients that do not change over sample period.
- Has good out-of-sample forecast.

Models of Volatility:

Many financial time series go through periods of tranquility and periods of volatility. Volatility and its measures are some of the most important concepts in finance.

Historical Volatility:

Measures the variance or standard deviation of returns over some period and uses it as measure of forecast of future volatility. It's used in option valuation and option pricing.

Characteristics of Financial Data:

- Non-stationary.
- Leptokurtosis: sharp peaks, fat tails.
- Volatility clustering.
- Leverage effect: Volatility rise more with price fall than price rise.
- Has coefficients that do not change over sample period.
- Has plenty noises as well as signals.
- Volatility is mean reverting.

Reading ACF and PACF

Non-stationary:

- If PACF has a single spike at $r_0 \approx 1$ and the rest is insignificant and ACF mostly significant and gradually decreasing $\Rightarrow$ non-stationary model.

(A)DF Test

- Failed to reject the null hypothesis: there is a unit root $\Rightarrow$ non-stationary.
- Reject the null hypothesis: stationary.

Test for White Noise

Box-Pierce Test

Null hypothesis: Residuals are White Noise.

$$Q = T \sum r^2 \tag{19}$$

Box-Ljung Test

Null hypothesis: Residuals are White Noise.

$$Q = T(T+2)\sum (T-K)^{-r^2} \tag{20}$$

Recall AR(2)

$$\begin{aligned}
Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} \\
\text{Regressing } \epsilon: & \\
b_i &= \alpha_1 b_{i-1} + \alpha_2 b_{i-2} \\
\text{We have:} & \\
\epsilon_{Y_{t+1}} &= \sum_{i=0}^{1} b_i \epsilon_{t-1} \\
&= b_0 \epsilon_t + b_1 \epsilon_{t-1} \\
&= \epsilon_t + \alpha_1 \epsilon_{t-1} \\
\text{Thus:} & \\
Var[Y_{t+1}] &= (1+\alpha_1^2)\sigma_\epsilon^2
\end{aligned} \tag{21}$$

Similarly:

$$\begin{aligned}
\epsilon_{Y_{t+2}} &= \sum_{i=0}^{2} b_i \epsilon_{t-1} \\
&= b_0 \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} \\
\text{Recall that: } b_i = \alpha_1 b_{i-1} + \alpha_2 b_{i-2} & \\
\text{We have:} \epsilon_{Y_{t+2}} &= \epsilon_t + \alpha_1 \epsilon_{t-1} + (\alpha_1 b_1 + \alpha_2 b_0)\epsilon_{t-2} \\
&= \epsilon_t + \alpha_1 \epsilon_{t-1} + (\alpha_1^2 + \alpha_2)\epsilon_{t-2} \\
\text{Thus:} & \\
Var[Y_{t+2}] &= \left[1 + \alpha_1^2 + (\alpha_1^2 + \alpha_2)^2\right]\sigma_\epsilon^2
\end{aligned} \tag{22}$$