IV. Time Series (Spring 2021)

When do you want to use time series?

1. Short-term forecasting

2. Information to estimate a structural model is not available

(Generally inferior to a structural model since cannot directly test structural hypotheses.)

   a. Theory is weak - Macroeconomics
   b. Observations on some variables are unavailable short-run price movements
      - daily
      - weekly

3. A dynamic model is needed - little dynamic theory.

4. A large number of observations are available (minimum 50-100)
   Why?
   Use data to select model, estimate a large number of parameters, no small sample properties due to lagged endogenous variables.

Want you to know

1. Stationarity, stability, and invertibility (relevant to dynamic simultaneous equation models)

2. Autoregressive integrated moving average (ARIMA)

3. Vector autoregression (VAR)

4. Granger Causality

5. Generalized autoregressive conditional heteroskedasticity (GARCH) models

6. Unit Root Tests

7. Cointegration

IV. a. <u>Basic Concepts</u>

Time Domain

Assume data are generated by some underlying stochastic process and attempt to model that process.
(Classes in stochastic processes STAT 5133)

(Regular Definition) A time series is data collected over time.

A time series is a random variable whose p.d.f. is a function of time.

$$f(Y,t;\theta) \quad Y=[Y_1,Y_2,\ldots,Y_T]'$$

Need to make some assumptions in order to model the process.

1. Assume Multivariate Normal

$$\underset{T\times 1}{Y} \sim N(\underset{T\times 1}{\mu}, \underset{T\times T}{\Sigma})$$

Have T means, T variances, $\frac{T(T-1)}{2}$ covariances
too many parameters to estimate

2. Assume Mean Stationarity

$$E(Y_t) = \mu \;\; ; t = 1, \ldots, T$$

now only have to estimate one mean

3. Assume Covariance (Autocovariance) Stationarity

$$R(h) = \text{cov}[Y_t, Y_{t+h}] = \text{cov}[Y_s, Y_{s+h}] \forall \; s, h, t$$

Note that this also implies that the variance is stationary since if $h = 0$,

$$\text{cov}(Y_t, Y_{t+0}) = \text{var}(Y_t) = \text{cov}(Y_s, Y_{s+0}) = \text{var}(Y_s)$$

now only have T+1 parameters; 1 mean 1 variance;
T-1 covariances
Still too many parameters

4. Assume covariance approaches zero as h becomes large (ergodicity)
(a) covariance can decay exponentially to zero
(b) covariance is truncated

Now, we estimate the p.d.f. with time series models.

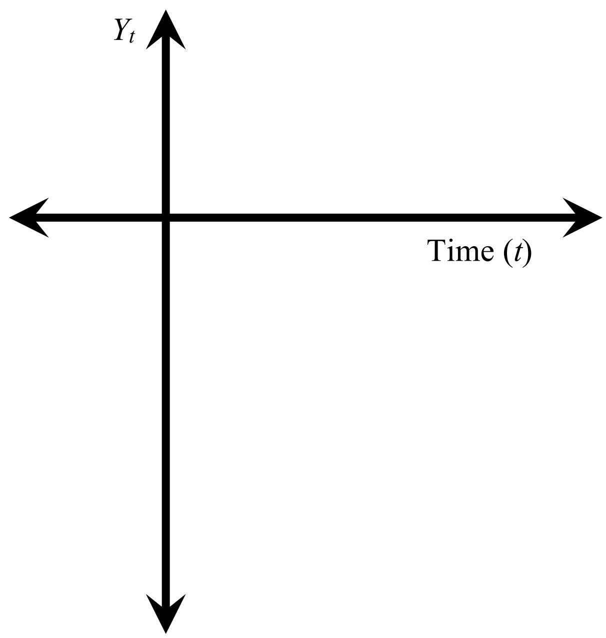
AR: $y_t = a y_{t-1} + e_t$
MA: $y_t = b e_{t-1} + e_t$

assume $a = b = 0.9$ and $e_t = \begin{cases} 1 \text{ if } t = 0 \\ 0 \text{ otherwise} \end{cases}$

$$Y_{-1} = 0$$

1. AR Model

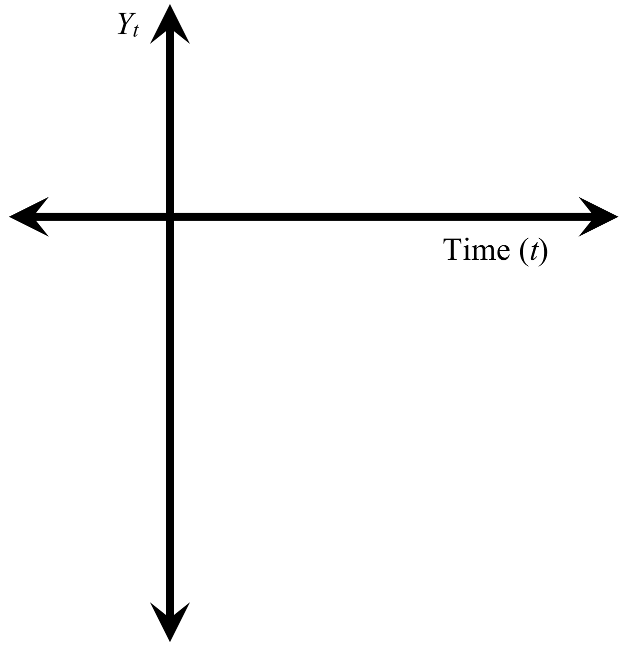
Time	Y_t
0	1
1	.9
2	.81
3	.729
4	.6561



Later values depend on what happened at time 0 (decay)

2. MA Model

Time	Y_t
0	1
1	.9
2	0
3	0
4	0



We could write an MA model equivalent to 1.

$$Y_t = .9e_{t-1} + .81e_{t-2} + .729e_{t-3} + .6561e_{t-3} + \cdots$$

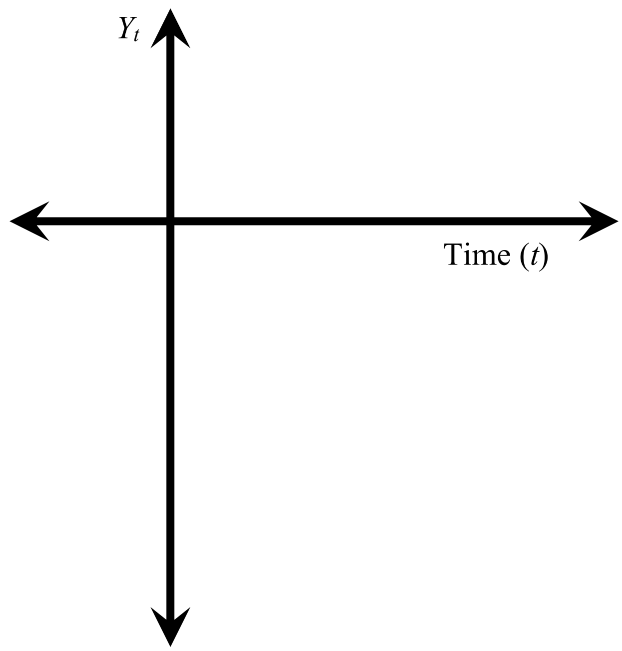
We can also write an AR model equivalent to 2:

$$Y_t = \sum_{i=1}^{\infty} (-1)^{(i-1)} b^i \, Y_{t-i} + e_t$$

Now what if $a = b = 1.1$

3. AR Model

Time	Y_t
0	1
1	1.1
2	1.21
3	1.331



Could we write this as
a MA model?

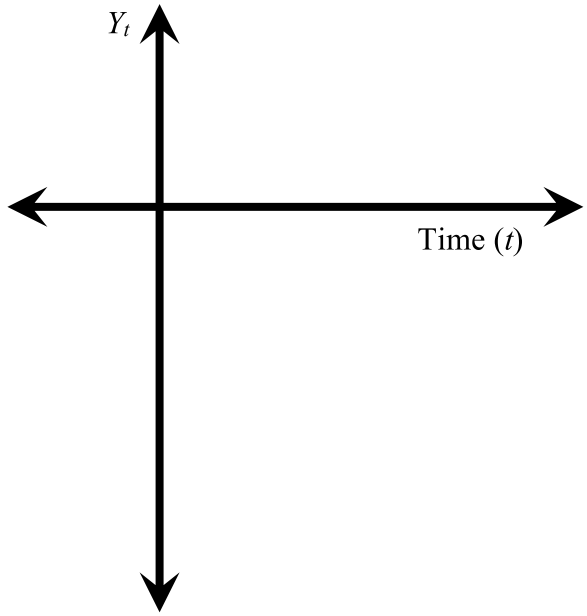
$$Y_t = \sum_{i=0}^{\infty} b^i e_{t-i}$$

$$= e_t + .1.1e_{t-1} + 1.21e_{t-2} + 1.331e_{t-3} + \cdots$$

No, we cannot because the parameters approach infinity implying nonstationary.

4. MA Model

Time	Y_t
0	1
1	1.1
2	0
3	0



Could we write this as an
AR Model?

$$Y_t = \sum_{i=1}^{\infty} (-1)^{(i-1)} b^i Y_{t-i} + e_t$$

$$= e_t + .1.1Y_{t-1} - 1.21Y_{t-2} + 1.331Y_{t-3} - \cdots$$

No, we cannot because the parameters approach infinity meaning not invertible.

$\therefore$ for first order processes $|a|<1$ if we can approximate
An AR(1) by an MA (∞)
If $|b|<1$ we can approximate an MA(1) by an AR(∞)

What about AR(p) or MA(q)?

Need eigenvalues.

Review of Complex Numbers

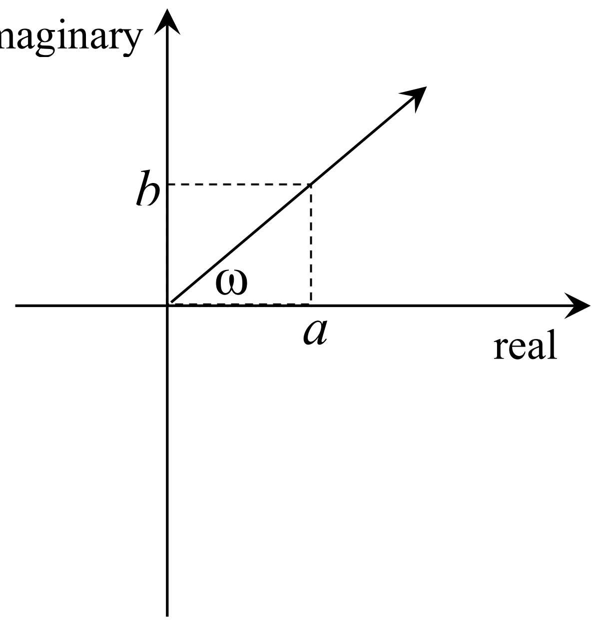
$i = \sqrt{-1}$

$\lambda = a + bi$ $\quad\quad$ a, b are real numbers

$\therefore$ complex number has an imaginary and a real part.
The modulus (absolute value is a special case)

$|\lambda| = \sqrt{a^2 + b^2}$

From trigonometry



$$\sin\omega = \frac{b}{|\lambda|}$$

$$\cos\omega = \frac{a}{|\lambda|}$$

$$\tan\omega = \frac{\sin}{\cos} = \frac{b}{a}$$

$$\therefore\ \omega = \tan^{-1}\left(\frac{b}{a}\right)$$

$\lambda=|\lambda|\,(\cos\omega+i\sin\omega)\rightarrow$will use this later

Consider where $|\lambda|=1$

$\cos\omega=a$
$\sin\omega=b$

unit circle is where $|\lambda|=1$

Complex conjugate ($\bar{\lambda}$)

$\lambda = a + bi$

$\bar{\lambda} = a - bi$

Characteristic Roots and Vectors $\quad Y_t = AY_{t-1} + CX_t$

Consider an n x n matrix **A**

$(\boldsymbol{A} - \lambda \boldsymbol{I})X = 0$

$\boldsymbol{I}$ is a n x n identity matrix

$\boldsymbol{X}$ is an n x 1 vector

λ is a constant

$$\text{Solution} \quad \begin{cases} \boldsymbol{X} = 0 & \text{not interesting} \\ \boldsymbol{X} \neq 0 & \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0 \end{cases}$$

$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$ is the characteristic equation of $\boldsymbol{A}$

Solve for λ to get a polynomial of degree n
Get n solutions that are called characteristic roots or eigenvalues

Properties

1. If $\boldsymbol{A}$ is symmetric, all the eigenvalues are real (Why SAS requires a symmetric matrix)

2. The roots of a positive (negative) definite symmetric matrix are all positive (negative)

3. The determinant of $\boldsymbol{A}$ is the product of the roots of $\boldsymbol{A}$

   If $\boldsymbol{A}$ is singular $\Rightarrow |\boldsymbol{A}| = 0 \Rightarrow$ at least one root of A is zero

   If $\boldsymbol{A}$ is nonsingular $\Rightarrow |\boldsymbol{A}| \neq 0 \Rightarrow$ none of the roots of $\boldsymbol{A}$ are zero

4. If roots are complex they appear in complex conjugate pairs:

   $\lambda = a \pm bi$ (that way the determinant will be a real number).

Example (Note discussion is in terms of time series models, but it also applies to reduced form econometric models.)

$Y_t = .5Y_{t\text{-}1} - .3Y_{t\text{-}2}$

$$\begin{bmatrix} Y_t \\ Y_{t\text{-}1} \end{bmatrix} = \begin{bmatrix} .5 & -.3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t\text{-}1} \\ Y_{t\text{-}2} \end{bmatrix}$$

$$\boldsymbol{A} - \lambda \boldsymbol{I} = \begin{bmatrix} .5 & -.3 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} .5 - \lambda & -.3 \\ 1 & -\lambda \end{bmatrix}$$

$(.5-\lambda)(-\lambda)+.3=0$

$\lambda^2-.5\lambda+.3=0$

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{+.5 \pm \sqrt{.25 - 4(1).3}}{2} = \frac{+.5 \pm \sqrt{-.95}}{2} = +.25 \pm .487i$$

$$|\lambda| = \sqrt{a^2 + b^2} = \sqrt{(.25)^2 + (.487)^2} = .547 \therefore \text{stable}$$

Complex roots $\therefore$ must be a cycle

What kind of cycle?

time domain $\Leftrightarrow$ frequency domain

Can write

$$\begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} S_1 & |\lambda_1|^t \cos(\omega t + \rho_1) \\ S_2 & |\lambda_2|^t \cos(\omega t + \rho_2) \end{bmatrix}$$

↖amplitude

Since $|\lambda|<1$ damped harmonic

$\omega = |\tan^{-1}\left(\frac{b}{a}\right)| = |\tan^{-1}\left(\frac{.487}{.25}\right)| = 1.0965$ radians

period $= \frac{2\pi}{\omega} = 5.73$ time units

Dynamic properties if λ_i is real

$0<\lambda_i<1$	stable damped
$-1<\lambda_i<0$	stable oscillatory
$\lambda_i>1$	unstable explosive
$\lambda_i<-1$	unstable oscillatory

Dominant root of a stable model is the one with the largest modulus (absolute value).

Definitions of stability & stationarity – equivalent conditions.

Stability (deterministic systems). A model is stable if for any initial values given the system it will reach an equilibrium.

Stationarity (stochastic systems). A time series is stationary if the joint distribution of any subset of observations does not change when a constant is added to the time subscript.

Autoregressive Process

$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + e_t$

zero mean since no intercept

AR(p) - autoregressive process of order p - p lags

$$e_t = \text{white noise} \begin{cases} E(e_t) & = 0 \\ E(e_t e_s) = \sigma^2 & \text{if } t = s \\ & 0 \text{ otherwise} \end{cases}$$

(Our usual assumptions give Gaussian white noise)

We can rewrite as

$Y_t^* = AY_{t-1} + E_t^*$ first-order linear homogenous

difference equation

$$\begin{bmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & & & \cdots & a_p \\ 1 & 0 & \cdots & & 0 \\ & \ddots & 0 & & \vdots \\ & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ \cdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

AR(*p*) is stationary if all roots of *A* are within the unit circle (i.e.,$|\lambda_i|<1 \; i=1,\ldots,p$) If we write as $A(L)Y_t = e_t$, stationary if roots of $A(L)$ are outside the unit circle. (Note that $L^k Y_t = Y_{t-k}$)

$Y_t=.9Y_{t-1} \Rightarrow (1-.9L)Y_t=0,\; 1-.9L = 0,\; L=\frac{1}{.9}$

Rewrite as an MA(∞)

$$Y_t^* = \sum_{i=0}^{\infty} A^i E_{t-i}^*$$

AR(*p*) = MA(∞) only if the *AR* process is stationary

Moving Average Process - assume zero mean

$Y_t = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \ldots + b_q e_{t-q}$

$Y_t = \sum_{i=0}^{q} b_i e_{t-i}$ MA(q) – q *lags*

Rewrite

$e_t = Y_t - \sum_{i=1}^{q} b_i e_{t-i}$

$E_t = \underline{Y}_t + BE_{t-1}$

$$\begin{bmatrix} e_t \\ \vdots \\ e_{t-q+1} \end{bmatrix} = \begin{bmatrix} Y_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} -b_1 & \cdots & & & -b_q \\ 1 & & & & 0 \\ 0 & \ddots & 0 & & \vdots \\ & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{bmatrix} \begin{bmatrix} e_{t-1} \\ \vdots \\ e_{t-q} \end{bmatrix}$$

Invertible if all roots of B are within the unit circle
or rewrite as $Y_t = B(L)e_t$
Invertible if all roots of $B(L)$ are outside the unit circle

If MA(q) is invertible then MA(q) = AR(∞)

A moving average process is always stationary.

Autoregressive Moving Average Model [ARMA(p,q)]

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=0}^{q} b_j e_{t-j}$$

ARMA is stationary if AR(p) portion is stationary

ARMA is invertible if MA(q) portion is invertible

$\therefore$ ARMA(p,q)=AR(∞)$\equiv$MA(∞) if the ARMA is

stationary and invertible

Often in practice an ARMA(p,q) process is approximated by a higher order AR process since AR is a simpler model.

IV. b. Box-Jenkins - type of time series modeling

Well developed only for univariate time series, but has been adapted for multivariate models.
Often works well for short-term forecasting.
Requires subjective decisions - whimsical.

Steps

1. Preliminary filtering ↖
2. Identification ←involve
3. Estimation pretesting
4. Diagnostic checking ↙

Only alternatives to pretesting

1. Give up
2. Arbitrarily select some model that may not be as close to the "true" model as the one selected using the data.

Filtering

Objective is to make the data stationary or remove deterministic components

Important since most economic data are not stationary: time trends, seasonality, changing variance. Often need to remove trend

1. Regress against a polynomial in time (trend stationarity)

2. Differencing - integrating
   I part of ARIMA(p,d,q)
   where d = degree of differencing

(First differences are usually preferred for price data.)

Tests for trends or unit roots

1. Autocorrelation function falls off very slowly
2. Dickey-Fuller unit root tests
3. Augmented Dickey-Fuller tests with time trend

Seasonality

1. Seasonal differencing – outdated
   $Y_t - Y_{t-s}$, s = the period of the seasonality
   12 if monthly data (now rarely used)
2. Include seasonal dummy variables
3. Use a set of sine and cosine functions

```
** Assume monthly data;
cos12=cos(2*3.1416*time/12);
sin12=sin(2*3.1416*time/12);
cos6=cos(2*3.1416*time/6);
sin6=sin(2*3.1416*time/6);
Proc Reg;
Model y = cos12 sin12 cos6 sin6;
```

4. Let the time series model capture seasonality - seasonal *AR* and *MA* terms (not recommended)

Changes in variance

1. Take logarithm of the data
2. Divide by estimated standard deviation (EGLS)
3. Autoregressive Conditional Heteroskedasticity (ARCH)

Identification [select orders p and q of the ARMA(p,q)]

The representation of an ARMA model is not unique (not a problem if you use strictly AR models)

Ex. $A(L)y_t = B(L)e_t$ multiply by $C(L)$

$$C(L)A(L)y_t = C(L)B(L)e_t$$

$$D(L)y_t = F(L)e_t$$

the objective of model building is to employ the smallest number of parameters which provide an adequate representation (parsimony)

Uses autocorrelation and partial autocorrelation functions

Autocorrelation

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{var}(Y_t)}$$

for an MA(q) $\rho_k = 0$ for $k>q$
for an AR(p) the autocorrelation dies out gradually
recall for an AR(1) $\Rightarrow$ $\rho\ \rho^2\ \rho^3\ \rho^4 \ldots$

Partial Autocorrelation
The partial autocorrelations of order j are the α_{jj} solved from

$$\rho_k = \sum_{i=1}^{j} \alpha_{ji}\, \rho_{k-i} \qquad k = 1, \ldots, j$$

The partial autocorrelation of degree j (α_{jj}) of an AR(p) is zero for $j > p$
Partial autocorrelations are the last AR coefficient from an AR(j) if model is stationary

The partial autocorrelation of a MA(q) dies out gradually.

Works well in identifying purely AR or MA models but becomes subjective in identifying an ARMA.

Initial identification is performed by examining the significance of partial autocorrelation and autocorrelation.

Approximate standard deviation is $\dfrac{1}{\sqrt{T}}$

(May suggest several alternative models)

Estimation

1. Nonlinear least squares - if MA, if AR - OLS
2. Maximum likelihood - (programs are available - of little concern to us. Use PROC ARIMA)

Note: MA sometimes uses backcasting

Diagnostic Checking

a. Delete insignificant parameters
b. Test residuals for white noise
c. Examine ACF and PACF of residuals

Either the model(s) is accepted or a new one(s) is proposed.

(The final model should have as few parameters as possible and white noise residuals)

White Noise Test

The objective of time series analysis is to filter the data down to white noise.

∴ we need tests to see if the modeling is complete

$e \sim WN$ if $\quad \mathrm{E}(e_t)=0 \qquad$ (zero mean)

$$\mathrm{E}(e_t e_s) = \begin{cases} \sigma^2 \ if \ t = s & \text{(homoskedastic)} \\ 0 \text{ otherwise} & \text{(no autocorrelation)} \end{cases}$$

Box-Pierce Q-statistic (Portmanteau test statistic) (tests H_0: no autocorrelation) often used with econometric models

$$Q(m) = T \sum_{k=1}^{m} \hat{r}_k^2 \xrightarrow{d} \chi^2(df)$$

T = no. of observations

$$\hat{r}_k^2 = \left(\frac{\operatorname{cov}(Y_t, Y_{t-k})}{\operatorname{var}(Y_t)} \right)$$

m is selected arbitrarily

for ordinary data df = m but, if data are residuals from a regression model, then df = m-p where p is the number of lags used in the regression equation

Ljung and Box have suggested a modification

$$Q(m) = T(T+2)\sum_{k=1}^{m} \hat{r}_k^2 \;/(T-K) \xrightarrow{d} \chi^2(df)$$

This is the standard since it is the most powerful.
same df adjustment for residuals

(Multivariate Portmanteau test is available, but is rarely used.)

Kennedy is negative about Box-Pierce and Ljung-Box tests, but Ljung-Box is what people use.

IV. c. Vector Autoregressive Models (easier than a multivariate ARMA)

- Make Y_t a vector rather than a scalar

Advantages

1. Unique representation - methods are objective
2. Easy to estimate - OLS
3. Easier to interpret than ARMA
4. Easier to identify - objective criteria
5. Limited pretesting - hypothesis tests still valid

Disadvantage

1. May not be a good approximation if the "true" model's moving average portion has a root near the unit circle.
2. Extra parameters, therefore less efficient & less powerful (poorer forecasts)

Steps

1. Typically a maximum possible order is selected supposedly using some prior knowledge
2. Minimization (maximization) of some criterion

Example:
Akaike's Information Criterion (AIC) (available in PROC STATESPACE, VARMAX, and ARIMA)

AIC = f(goodness of fit [reward]) + g(no. of parameters
= reward + penalty

Tradeoff between accuracy and parsimony

$$AIC(m) = \ln|\tilde{\Sigma}_m| + 2d^2 m / T \quad m = 0, 1, \ldots, \overline{p}$$

where

m is the order of the model

T is the number of observations
d is the number of variables (dimensionality of the time series)
$|\tilde{\Sigma}_m|$ is the determinant of the prediction error

covariance matrix

$$\tilde{\Sigma}_m = (Y - X\ \hat{\beta})'(Y - X\hat{\beta})/T$$

$$d\,x\,d \qquad T\,x\,d \quad T\,x\,k \quad k\,x\,d$$

The order selected is the value of *m* for which the AIC is minimized.

The AIC tends to overestimate the "true" order. Corrected AIC is now often used.

Numerous other methods have been suggested.

One alternative is an F-test (encompassing principle).

Standard pretesting

$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + A_3 Y_{t-3} + \ldots + A_n Y_{t-n}$

H_0: $A_n = 0$, if reject select order n
Else (restrict $A_n = 0$) test H_0: $A_{n-1} = 0$ if reject select order n - 1,
Else restrict $A_{n-1} = 0$ and test H_0: $A_{n-2} = 0$

Schwarz Criterion (SBC [Bayesian]) in PROC ARIMA
- available in PROC RSQUARE currently popular

$$SBC(m) = \ln\left|\tilde{\Sigma}_m\right| + K_m \ln T / T \quad m = 0, 1, \ldots, \bar{p}$$

where K_m is the total number of parameters ($d^2 m$) for a VAR

Use PROC STATESPACE to select the order (AIC)

Use PROC REG or PROC SYSLIN to estimate

Or use VARMAX to do both

IV. d. <u>Granger Causality</u>

A variable X causes a variable Y, with respect to a given set of information that includes both X and Y, if Y can be predicted more accurately using past values of X than if the information about X is not used.

- Not the definition used by statisticians or philosophers. It equates correlation and causality.

Undirectional causality

Y causes X

X does not cause Y

Bidirectional causality (feedback)

Y causes X

X causes Y

Indirect causality

Y causes Z

Y does not cause X

Z causes X

Instantaneous causality

Cannot determine the direction

Many ways to test for Granger causality, the most popular way is:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} a_{11}(i) a_{12}(i) \\ a_{21}(i) a_{22}(i) \end{bmatrix} \begin{bmatrix} Y_{1t-i} \\ Y_{2t-i} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

H_0: Y_2 does not cause $Y_1 \Rightarrow a_{12}(i) = 0$, $i = 1, ..., p$

H_0: Y_1 does not cause $Y_2 \Rightarrow a_{21}(i) = 0$, $i = 1, ..., p$

$\therefore$ can use a Wald, LR, or LM test

In SAS

```
Input Y X;
Y1 = LAG(Y); Y2 = LAG2(Y); Y3 = LAG3(Y);
X1 = LAG(X); X2 = LAG2(X); X3 = LAG3 (X);

PROC REG;
MODEL Y = Y1 Y2 Y3 X1 X2 X3;
TEST X1 X2 X3;
MODEL X = Y1 Y2 Y3 X1 X2 X3;
TEST Y1 Y2 Y3;
```

Time Series vs. Structural Econometric Model -

(Zellner & Palm *J. Econometrics* 1974)

(May include in 2019)

Although time series and econometric models are derived from quite different assumptions, they are really quite similar. Time series model is a reduced form.

Structural Econometric Model

$\Gamma Y + BX + E = 0$

reduced form

$Y = \pi X + U$

$$Y_t = \sum_{i=1}^{p} \tau_i Y_{t-i} + \sum_{j=0}^{n} \lambda_j X_{t-j} + U_t \text{ Dynamic SEM}$$

How is this different from a VAR? - no equations for X - and no structural restrictions what if $d(\mathrm{L})X_t = V_t$

$$\begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} U_t \\ V_t \end{bmatrix}$$

Since X_t is exogenous $c(L) = 0$

Granger causality has been proposed as a test of exogeneity

A time series model is equivalent to an unrestricted reduced form econometric model.

IV. e. General Autoregressive Conditional Heteroskedasticity (GARCH)

ARCH (Autoregressive Conditional Heteroskedasticity)

Engle *Econometrica* 1982 pp. 987-1007

Let the variance follow a time series process.

Model

$Y_t = X_t \delta + \varepsilon_t$

$\varepsilon_t | \theta_{t-1} \sim N(0, h_t)$

where θ_{t-1} is information available in time $t-1$

$$h_t = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2$$

Essentially autoregressive in ε_t^2

$$\varepsilon_t^2 = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + v_t$$

$$h_t = \mathrm{var}(\varepsilon_t \mid \theta_{t-1}) = E(\varepsilon_t^2 \mid \theta_{t-1}) = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2$$

GARCH - Bollerslev *J. Econometrics* 1986
pp. 307-327

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

ARMA in ε_t^2

$$\varepsilon_t^2 = c + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j v_{t-j} + v_t$$

interpretation of coefficients is different

$a_i = \alpha_0 + \beta_i \quad b_i = \beta_i$

can be extended to multivariate GARCH

can add exogenous variables to variance equation

GARCH (1, 1) is the most common

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

$\alpha_1 \geq 0$, $\beta_1 \geq 0$ $\alpha_1 + \beta_1 < 1$

for positive variance and covariance stationarity

Lagrange Multiplier test for ARCH (Same as

McGuirk dynamic heteroskedasticity test)

```
PROC AUTOREG;
MODEL Y = X / ARCHTEST;
```

1. Estimate assuming no heteroskedasticity (restricted model).

$Y_t = X_t \delta + \varepsilon_t \qquad t = 1, \ldots, T$

run OLS

$$\hat{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \hat{\varepsilon}_{t-i}^2$$ often use $q = 1$

use F-test

McLeod-Li test (Ljung-Box test with) $\hat{\varepsilon}_t^2$

$$\rho_k^* = \frac{\mathrm{cov}(\hat{\varepsilon}_t^2, \hat{\varepsilon}_{t-k}^2)}{\mathrm{var}(\hat{\varepsilon}_t^2)}$$

$$Q_{(m)}^* = T(T+2)\sum_{k=1}^{m} \rho_k^{*2} / (T-k) \xrightarrow{d} \chi^2(\mathrm{df})$$

Under H_0: no conditional heteroskedasticity

Now consider the GARCH (1,1)

$Y_t = \beta' X_t + \varepsilon_t$, $\varepsilon_t \sim N(0, h_t)$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

EGLS estimation

1. Estimate

   $Y_t = X_t\,\delta + e_t$ $\quad t=1, \ldots, T \quad$ by OLS

   Obtain $\hat{\varepsilon}_t$ consistent estimates

2. Estimate ARMA (PROC ARIMA)

$$\varepsilon_t^2 = c + a\varepsilon_{t-1}^2 + bv_{t-1} + v_t$$

   $a = \alpha + \beta \quad b = -\beta \quad$ Consistent estimates

3. Use EGLS

   $\hat{\delta}_{EGLS} = (X'\hat{\Phi}^{-1}X)^{-1}X'\hat{\Phi}^{-1}Y$

$$\hat{\Phi} = \begin{bmatrix} \hat{\varepsilon}_1^2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \hat{\varepsilon}_{T^2} \end{bmatrix}$$

Asymptotically efficient estimates

Maximum Likelihood Estimation

assume normality

Judge, et al. p. 539

$y_t = x_t\,\delta + \varepsilon_t \quad t = 1, \ldots, T$

$\varepsilon_t | \Phi_{t\text{-}1} \sim \mathrm{N}(0, h_t)$

exogenous variables

$h_t = \alpha_0 + \alpha_1\,\varepsilon_{t\text{-}1} + \beta\,h_{t\text{-}1} + \theta Z_t$

log likelihood for multivariate normal

$\mathrm{L}(\delta, \alpha_0, \alpha_1, \beta, \theta \mid y, X, Z)$

$$= -\frac{T}{2}\ln 2\pi - \frac{1}{2}\ln|\Phi| - \frac{1}{2}(y - X\delta)^{/}\,\Phi^{-1}(y - X\delta)$$
$$= -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln h_t - \frac{1}{2}\sum_{t=1}^{T}(y_t - X_t)^2/h_t$$

in SAS
PROC AUTOREG;
MODEL *Y* = time/nlag = 2 garch = (*q*=1, *p*=1) maxit=50;

IV. f. Unit root tests

what is a unit root? $I(1)$

ARIMA (p, d, q)

$AR(p)$ MA(q) $I(d)$

$A(L)y_t$

if $A(L) = 1 - L$

root (where A(L)=0) is $L = 1$ i.e. unit root

y_t is nonstationary

(infinite variance ∴ Gauss-Markov does not hold)

Augmented Dickey-Fuller Test

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \sum_{j=1} \delta_j \Delta Y_{t-j} + \varepsilon_t$$

H_0: Y_{t-1} has a unit root $\Rightarrow \alpha_1 = 0$

H_a: Y_{t-1} does not have a unit root $\Rightarrow \alpha_1 < 0$

The usual t-statistic does not have a t-distribution, so it is called a pseudo t-statistic. Dickey and Fuller (1981) have tabulated the distribution of the pseudo t-statistic.

Variations

Can also test $\alpha_0 = 0$ no concern to us

Dickey-Fuller test $p = 0$ (no lagged differences)

Test unit root vs. Quadratic trend

non-nested

nest in a more general model

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 t + \sum_{j=1}^{p} \delta_j \Delta Y_{t-j} + \varepsilon_t$$

H_0: Y_t has a unit root : $\alpha_1 = 0$

H_A: Y_t does not have a unit root : $\alpha_1 < 0$

Phillips-Perron test

nonparametric correction for autocorrelation

KPSS Test : Null hypothesis is stationarity around a linear trend lots of other unit root tests available-ADF is still most common

Structural breaks are often tested.

intercept shifters

slope shifter on time trend - Sims, Stock, and Watson *Econometrica* Jan. 1990

Panel unit root tests are becoming common-available in STATA

IV. g. <u>Cointegration</u>

long-run equilibrium

- observation period length is important
- observation frequency is unimportant

two series are cointegrated if both series are $I(1)$ but a linear function of the series is $I(0)$

Example

$y_{1t} \sim I(1)$, $y_{2t} \sim I(1)$

$z_t = cy_t = y_{1t} - 2y_{2t}$

$z_t \sim I(0)$

$c = [1, -2]$ is the cointegrating vector

$y_t \sim C\ I(1,1)$

important, can run regressions on levels if cointegrated

$$y_{1t} = a + by_{2t} + \varepsilon_t$$

var (ε_t) is finite if y_{1t} and y_{2t} are cointegrated

$\therefore$ GAUSS-MARKOV theorem holds

Why is it a big deal in more than time series analysis?

To test for cointegration

1. determine if y_{1t} and y_{2t} are both I(1)

   • test H_0: unit root, fail to reject

   • H_0: two unit roots, reject $\therefore$ not I(2)

2. regress

   $y_{1t} = \beta_0 + \beta_1 y_{2t} + u_t$

   obtain residuals

   choice of dependent variable is arbitrary

3. Determine if u_t has a unit root using ADF:

$$\Delta\hat{u}_t = \alpha\,\hat{u}_{t-1} + \sum_{j-1}^{p} \varnothing_j\, \Delta\hat{u}_{t-j} + v_t$$

Numerous alternative tests.

IV. h. <u>Error Correction Models</u>

The error correction model is

$$\underset{2x1}{\Delta Y_t} = \underset{2x1}{\rho} \underset{1x1}{z_{t\text{-}1}} + \sum_{j=1}^{p} \underset{2x2}{\delta_j} \underset{2x1}{\Delta Y_{t\text{-}j}} + \underset{2\times 1}{\varepsilon_t}$$

$$Y_t = \begin{bmatrix} y_{1\mathrm{t}} \\ y_{2\mathrm{t}} \end{bmatrix} \qquad \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$ similar to Dickey-Fuller regression

$z_{t\text{-}1} = cY_{t\text{-}1} = y_{1t\text{-}1} - ay_{2t\text{-}1}$

$|\rho_1| + |\rho_2| \neq 0$

as y_1 and y_2 get further apart the error correction term gets larger

$$\Delta y_{1t} = \rho_1 (y_{1t\text{-}1} - \alpha y_{2t\text{-}1}) + \sum_{j=1}^{p} \delta_{1j}\, \Delta y_{t\text{-}j} + \varepsilon_{1t}$$

$$\Delta y_{2t} = \rho_2 (y_{1t\text{-}1} - \alpha y_{2t\text{-}1}) + \sum_{j=1}^{p} \delta_{2j}\, \Delta y_{t\text{-}j} + \varepsilon_{2t}$$

rewrite as

$$\text{(A)}\ \Delta y_{1t} = \rho_1 y_{1t\text{-}1} - \rho_1 \alpha y_{2t\text{-}1} + \sum_{j=1}^{p} \delta_{1j}\, \Delta y_{t\text{-}j} + \varepsilon_{1t}$$

$$\text{(B)}\ \Delta y_{2t} = \rho_2 y_{1t\text{-}1} - \rho_2 \alpha y_{2t\text{-}1} + \sum_{j=1}^{p} \delta_{2j}\, \Delta y_{t\text{-}j} + \varepsilon_{2t}$$

The ECM has a nonlinear restriction

Could estimate (A) and (B) directly with PROC MODEL

Engle and Granger recommend a 2-step estimation method

Step 1 estimate the cointegrating regression

$y_{1t} = \alpha y_{2t} + z_t$

$\hat{z}_t$ are residuals from the above regression

$\hat{z}_t = y_{1t} - \hat{\alpha} y_{2t}$

Step 2:

regress

$$\Delta y_{1t} = \rho_1 \hat{z}_t + \sum_{j=1}^{p} \delta_{1j} \Delta y_{t-j} + \varepsilon_{1t}$$

$$\Delta y_{2t} = \rho_2 \hat{z}_t + \sum_{j=1}^{p} \delta_{2j} \Delta y_{t-j} + \varepsilon_{2t}$$

estimates of parameters and their standard errors are consistent and asymptotically efficient

(no generated regressor problem in this case)

Johansen's Cointegration

Most common method of testing cointegration.

Useful for more than two time series

likelihood ratio test

Model

$x_t = \pi_1 x_{t-1} + \ldots + \pi_k x_{t-k} + \varepsilon_t \quad t = 1, 2, \ldots$

ε_t - p-dimensional Gaussian r.v. with mean zero, variance

x_t can be nonstationary

$A(Z) = I - \pi_1 Z - \ldots - \pi_k Z^k$

unit root $\Rightarrow$ $|A(Z)|$ has roots at $z = 1$

$A(Z)|_{z=1} = \pi = \mathrm{I} - \pi_1 - \ldots - \pi_k$

has rank $r < p$ (p time series, each with a unit root, but some are linearly dependent)

$\underset{pxp}{\pi} = \underset{pxr}{\alpha} \ \underset{rxp}{\beta'}$

β - cointegration vectors estimate rank of $\alpha\beta'$

outline of procedures

regress Δx_t and x_{t-k} on lagged differences

get residuals - calculate product moments ($\hat{\varepsilon}_i' \ \hat{\varepsilon}_j$)

corrected for the lagged differences

LR test is a function of eigenvalues of the product moment matrix's asymptotic distribution. It is an integral of multivariate Brownian motion-stochastic integral

Regress

$$\Delta x_t = A_1 \Delta x_{t-1} + \ldots + A_{k-1} \Delta x_{t-k+1} + R_{0t}$$

$$x_{t-k} = B_1 \Delta x_{t-1} + \ldots + B_{k-1} \Delta x_{t-k+1} + R_{kt}$$

$$\underset{(pxp)}{S_{ij}} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt} \quad i, j = 0, k \text{ (only 0 and } k\text{)} \quad (8)$$

Find the ordered eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p$ of $S_{k0} S_{00}^{-1} S_{0k}$ w.r.t. s_{kk}

which is the solutions to

$$|\lambda s_{kk} - s_{k0} s_{00}^{-1} s_{0k}| = 0 \qquad (9)$$

in practice

$$\hat{\lambda} = \text{EIGVAL}\left(\text{CHOL}(s_{kk})^{-1} s_{k0}\, s_{00}^{-1}\, s_{0k}(\text{CHOL}(s_{kk}^{-1}))\right)$$

H_0: at most r cointegration vectors

$$LR = -T \sum_{i=r+1}^{p} \ln(1-\hat{\lambda}_i) \qquad \text{(TRACE TEST)(2)}$$

where $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ are the $p - r$ smallest

eigenvalues (squared canononical correlations between x_{t-k} and Δx_t)

MAX EIGENVALUE TEST

$$-T \ln(1-\hat{\lambda}_{r+1})$$

Use Table A3 in Johansen and Juselius (1990)

Many new variations available in R
Structual VAR
Threshold VAR
Threshold cointegration
Multivariate GARCH – volatility spillover

Time Series Analysis

Frequency Domain Spectral Analysis spectrum
Review of trigonometry light
consider the right triangle prism

wavelength

$$\sin(\theta) = \frac{a}{b} \quad \cos(\theta) = \frac{a}{b} \qquad \text{tangent}(\theta) = \frac{a}{b}$$

θ is sometimes measured in degrees, but we will use radians since that is what computers use

radians = 180^{o}

sin(0) = 0 cos(0) = 1

Now consider the unit circle

Example:

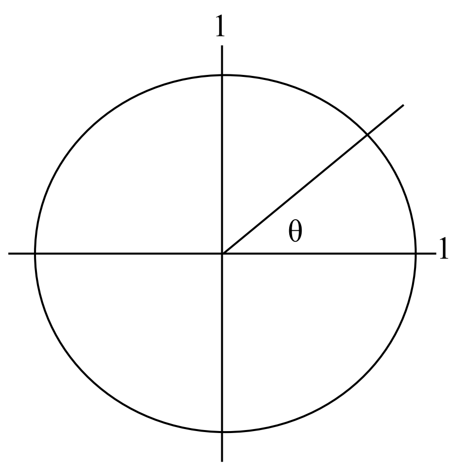
$\theta = 45^{o} = \dfrac{\pi}{4}\, rad.$

c is always 1 $\quad\quad \therefore\ a$ = sine $\quad b$ = cosine

$$\sqrt{a^2 + b^2} = 1 \quad \text{if } a = b$$

$$a = b = \sqrt{.5} = .7071$$

Graphically



Follows a regular cyclical pattern

Definitions

- Amplitude: the maximum height of the cycle
- Phase Shift: the distance the maximum height from the origin
- Frequency: the portion of a cycle completed within one time period
- Period: the length of the cycle (number of time periods it takes to complete one cycle)

How to mathematically represent this function sinusoid

$f(\theta) = a\cos(\theta) + b\ \sin(\theta)$

$\quad = c\cos(\theta\text{-}\rho)$

where

$$c = \sqrt{a^2 + b^2} = \mathit{amplitude} \quad \rho = \tan^{-1}\left(\frac{b}{a}\right) = \mathit{phase\ shift}$$

if $\theta = \omega t$ where t is time

ω is a constant measured in radians

then length of the cycle = period = $\dfrac{2}{\omega}$

frequency = $\dfrac{1}{\mathit{period}} = \dfrac{\omega}{2} =$ # of cycles/time period

Frequency domain is based on Fourier Series

First, assume the mean, variance, and autocovariance of Y_t are not a function of time (i.e., Y_t is a stationary process, Y_t has a mean of zero).

Fourier Series

$$w_j = \frac{2}{N} \cdot j \quad j=1, \ldots, \frac{N}{2}$$

$$\text{or } Y_t = \sum_{j=1}^{\frac{N}{2}} C_j \cos(w_j t + p_j)$$

No error term, since no degrees of freedom, get a perfect fit

Thus, any time series can be written as the sum of $\frac{N}{2}$ sinusoids.

What are the periods of the sinusoids used?

$$\frac{w_j}{2\pi} = \text{frequency} \quad \frac{2\pi}{w_j} = \text{period} \quad w_j = \frac{2\pi}{N} \cdot j \quad j=1, \ldots, \frac{N}{2}$$

$$\therefore \textit{period}_j = \frac{2\pi}{\left(\frac{2\pi}{N} \cdot j\right)} = \frac{N}{j} \quad j = 1,, \cdots, \frac{N}{2}$$

Example
if $N = 64$
then periods used will be

$$64, \frac{64}{2}, \frac{64}{3}, \frac{64}{4}, \ldots, \frac{64}{32}$$

$\therefore$ more small cycles are used
a_j and b_j are parameters

Harmonic Analysis is a regression on the Fourier Series - a parametric approach

Spectral Analysis - a non-parametric approach
  spectral density measure the contribution of the jth frequency to the variance of Y_t

Part of spectral analysis we will be concerned with is the periodogram. Also gain, coherence, and phase shift.

Periodogram - graph of estimates of the spectral density function

Can calculate estimates two ways

1. Estimate a_j and b_j by regression then amplitude measures the strength of the cycle for that frequency

$$f(w_j) = \frac{NC_j^2}{4\pi} = \sqrt{a_j^2 + b_j^2} \cdot \frac{\frac{N}{2}}{4\pi}$$ Sometimes

c_j^2 *vs.* w_j

on $\ln(c_j)$ *vs* $w_j \leftarrow$ squeezes

2. Estimate from the autocovariance function
   Autocovariance - auto means self

$$\hat{\gamma}_k = \hat{\gamma}_{-k} = \sum_{t=k+1}^{N} \frac{(y_t - \bar{y})(y_{t-k} - \bar{y})}{N-K} \quad (-N+1 \le k \le N-1)$$

only N-k values because of lags

spectral density

$$f(w_j) = \frac{1}{\pi} \sum_{k=N+1}^{N-1} \gamma_k \cos(w_j k)$$

$\therefore$ can estimate either with $\hat{\gamma}_k$'s or $\hat{a}_j$ and $\hat{b}_j$'s

Review of trigonometry

(Why important -
frequency $\Leftrightarrow$ time - *AR* models follow regular cyclical patterns

Why are sine and cosine functions useful in explaining cyclical behavior)

Consider the right triangle

$c = \sqrt{a^2 = b^2}$ hypotenuse

sine	cosine	tangent	
OH	AH	OA	Oscar Had a Hunk of Apple

$\sin\theta = \frac{a}{c} \quad \cos\theta = \frac{b}{c} \quad \tan\theta = \frac{a}{b}$

θ is sometimes measured in degrees, but on a computer you must use radians

$\sin(0) = 0$ $\cos(0) = 1$ $180° = \pi$ radians

Now consider the unit circle

circumference = $2\pi r = 2\pi$ 360° degrees
$\cos(\theta + 2\pi) = \cos\theta$

say θ is 45° same as $\frac{\pi}{4}$ radians

what are sine and cosine

c is always 1 $\therefore$ a=sine b=cosine

$$\sqrt{a^2 + b^2} = 1 \quad \text{if } a = b \quad a^2 = b^2 = .5 \quad \sqrt{.5} = .7071$$

range $-1 \le \sin\theta \le 1$ $-1 \le \cos\theta \le 1$

graph

sinusoid $\rightarrow$ $a \cos \theta + b \sin \theta$

need both sine and cosine to get the phase shift

to get phase shift of π let $a = -1$ $\quad$ $b = 0$

to get a phase shift of $\frac{\pi}{4}$ let $a = .7071$ $\quad$ $b = .7071$

example - use a sinusoid to approximate seasonality in monthly data period = 12   frequency = $\frac{1}{12}$

$a_i \cos(w_i t) + b_i \cos(w_i t)$

$\frac{w_i}{2\pi}$ = frequency $\frac{w_i}{2\pi} = \frac{1}{12}$ $w_i = \frac{2\pi}{12}$

$$Y_t = a\cos\left(\frac{2\pi}{12}t\right) + b\cos\left(\frac{2\pi}{12}t\right) \quad t = 1, \cdots, N$$

a, b are parameters to estimate
sinusoids are orthogonal to each other

$\therefore$ estimates of a, b would not change if other sinusoids were added to the regression equation

Fourier Series

$$Y_t = \sum_{j=1}^{\frac{N}{2}} a_j \cos w_j t + b_j \sin w_j t + \overline{Y}$$

$w_j = \frac{2\pi}{N} \cdot j$ $\therefore$ the frequencies considered are the smallest frequencies that fit the dataset

perfectly

graphically

follow a regular cyclical pattern

(use →) sinusoid: the sum of a sine and cosine function

$f(\theta) = a\cos(\theta) + b\sin(\theta)$

through trig.

$a\cos\theta + b\sin\theta = C\cos(\theta - p)$

$C = \sqrt{a^2 + b^2} \rightarrow amplitude$

$p = \tan^{-1}\left(\frac{b}{a}\right) \rightarrow phase\ shift$

if $\theta = wt$ where t is time
w is a constant measured in radians

then length of cycle is $\frac{2\pi}{w}$

frequency

example if N = 64

then periods of 64, $\frac{64}{2}$, $\frac{64}{3}$, $\frac{64}{4}$, ..., $\frac{64}{32}$ will be used

∴ more small cycles are considered

a_j and b_j are parameters to be estimated

Spectral Density

$$f(w_j) = \frac{1}{\pi} \sum_{k=N+1}^{N-1} \gamma_k \cos w_j k$$

$$= \frac{1}{2}(a_j^2 + b_j^2)\frac{N}{2\pi} \quad w_j = \frac{2\pi j}{N} \; j = 1, \cdots, \frac{N}{2}$$

where $\gamma_k = \mathrm{Cov}(Y_t, Y_{t-k})$ $\quad -N+1 \leq k \leq N-1$

estimate $\hat{\gamma}_k$ or $\hat{a}_j$, $\hat{b}_j$ to get estimates of the spectral density

periodogram: graph of the estimated spectral density function

problem is that these estimates are unbiased but inconsistent, because the variance does not converge to zero as sample size increases