MATH 681

TIME SERIES ANALYSIS

Dr David A Stephens
Burnside 1235
d.stephens@math.mcgill.ca
www.math.mcgill.ca/~dstephens/681/

Objectives of Time Series Analysis :

- modelling for sequences of random variables indexed by time

$$X_1, X_2, \ldots, X_t, \ldots.$$

  - joint pdf
  - marginal pdf
  - first and second order moment structure

    i.e mean, variance

    covariance

Typically $X_1, X_2, \ldots, X_t$ ($\{X_t\}$) will not be independent

- **Forecasting** (Prediction)

Given observed values $\{x_t\}$ of $\{X_t\}$

$$X_1 = x_1$$
$$X_2 = x_2$$
$$\vdots$$
$$X_T = x_T$$

forecast the "future" values of

$$X_{T+1}, X_{T+2}, \ldots, X_{T+k}$$

- expected values
- pdfs / jointpdf

1.1 Simple Stationary Time Series Models

DEFINITION A time series is a sequence of random variables

$$\{X_t\}_{t=1,\ldots,T} = (X_1, \ldots, X_T)$$

or a realization (observed value)

$$\{x_t\}_{t=1,\ldots,T} = (x_1, \ldots, x_T)$$

of these variables.

A time series model is a probability model specifying either

(a) the joint pdf of $(X_1, \ldots, X_T)$

or (b) the first and second order moments of $(X_1, \ldots, X_T)$ (mean, covariance structure)

The usual joint cdf for $(X_1, \ldots X_T)$ is given by the probability

$$P\left[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_T \leq x_T\right]$$

This function of arguments $(x_1, \ldots, x_T)$ is usually written

$$F_{X_1, X_2, \ldots, X_T}(x_1, x_2, \ldots, x_T)$$

and fully specifies the joint probability structure of $(X_1, \ldots, X_T)$,

INCLUDING MOMENTS

SPECIAL CASE : INDEPENDENCE

If $X_1, \ldots, X_T$ are <u>independent</u> then

$$P[X_1 \le x_1, X_2 \le x_2, \ldots, X_T \le x_T]$$

$$= \prod_{t=1}^{T} P[X_t \le x_t]$$

or

$$P[X_t \le x_t \mid X_1 = x_1, \ldots, X_{t-1} = x_{t-1}]$$

$$= P[X_t \le x_t]$$

i.e the observed values of $X_1, \ldots, X_{t-1}$ contain no information about X_t

[no forecasting benefit]

Often, in time series modelling, attention is focussed on <u>moments</u>

$E[X_t]$ $t=1,..,T$ expectation

$Var[X_t]$ $t=1,..,T$ variance

$Cov[X_t, X_s]$ $t \neq s$ covariance

rather than the full joint distribution.

These are the "second-order properties" of $(X_1, .., X_T)$.

Typically $X_1, \ldots, X_T$ will be of the same type (discrete or continuous), as the represent the same feature measured over successive days/weeks/months etc.

EXAMPLE IID PROCESS $\{X_t\}$

$$P[X_t = 1] = p$$

$$P[X_t = -1] = 1-p$$

EXAMPLE "RANDOM WALK"

$\{X_t\}$ as for the iid process above

$\{S_t\}$ defined by

$$S_t = \sum_{i=1}^{t} X_i$$

Note that, if $p = \frac{1}{2}$

$$E[X_t] = 0$$

$$Var[X_t] = 1$$

$$E[S_t] = 0$$

$$Var[S_t] = t$$

so both processes $\{X_t\}$ and $\{S_t\}$ are <u>zero mean</u> processes in this case.

Note also that

$$X_t = S_t - S_{t-1}$$ "DIFFERENCING"

(if $S_0 \equiv 0$)

If $p \neq 1/2$

$$E[X_t] = 2p - 1$$

$$E[S_t] = (2p-1)t$$

If $p < 1/2$, $\{S_t\}$ has <u>downward</u> drift

$p > 1/2$, $\{S_t\}$ has <u>upward</u> drift

Note that whatever the value of p,

$$\mathrm{Var}[S_t]$$

grows linearly with t.

1-2.

MODELS WITH TRENDS AND SEASONALITY

TRENDS For many real time series data sets, it is evident that a deterministic trend is present

Thus a model of the form

$$X_t = m_t + Y_t$$

where m_t is a deterministic function of t

eg $m_t = a + bt$

may be useful, where

$\{Y_t\}$ is a suitably chosen sequence of r.v.s.

If $\{Y_t\}$ is a zero-mean process then the function m_t can be estimated using least-squares

That is, if $x_1, \ldots, x_T$ are observed, and it is believed that

$$m_t = a_0 + a_1 t + a_2 t^2 \equiv m_t(\underset{\sim}{a})$$

we may estimate (a_0, a_1, a_2) using Ordinary Least Squares (OLS)

i.e.

$$\hat{\underset{\sim}{a}} = \underset{\underset{\sim}{a}}{\operatorname{argmin}} \sum_{t=1}^{T} \left(x_t - m_t(\underset{\sim}{a})\right)^2$$

[CONSISTENT IF NOT EFFICIENT!]

EXAMPLE LAKE HURON DATA (B+D Ex 1.3.5)

X_t - level (in feet) of Lake Huron
1875 - 1972 ($T = 98$)

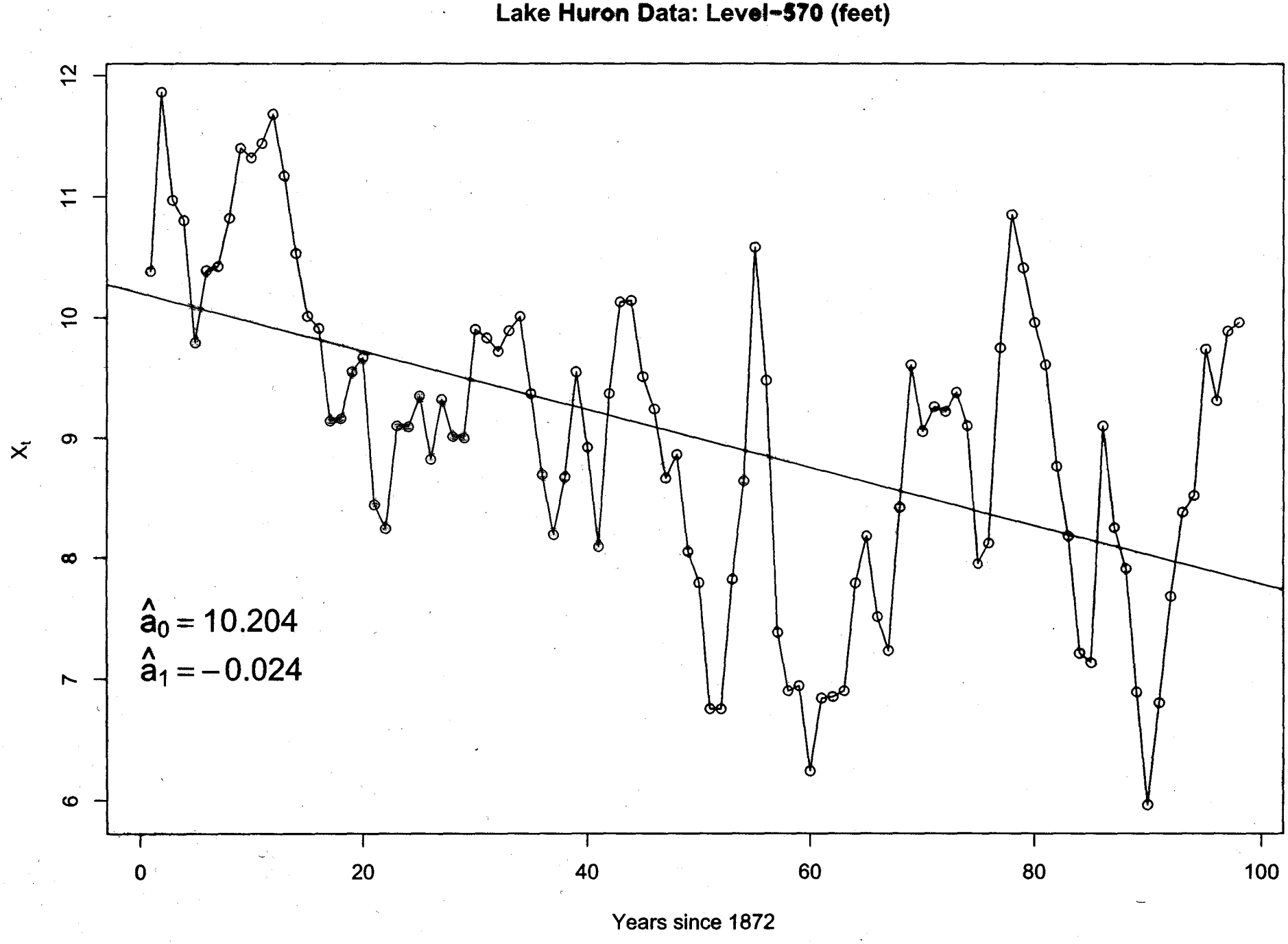
Suppose

$$X_t = a_0 + a_1 t + Y_t \qquad t = 1, \ldots, 98$$

Find $(\hat{a}_0, \hat{a}_1)$ by minimizing

$$\sum_{t=1}^{98} \left(x_t - a_0 - a_1 t \right)^2$$

	est.	(a.s.e)
$\hat{a}_0$ =	10.202	(0.230)
$\hat{a}_1$ =	-0.0242	(0.004)


Lake Huron Data: Level-570 (feet)
X_t
12
11
10
9
8
7
6
0
20
40
60
80
100
Years since 1872
$\hat{a}_0 = 10.204$
$\hat{a}_1 = -0.024$

Estimate $\{Y_t\}$ $t = 1, \ldots, 98$ by

$$\hat{y}_t = x_t - \hat{m}_t$$

$$= x_t - \hat{a}_0 - \hat{a}_1 t$$

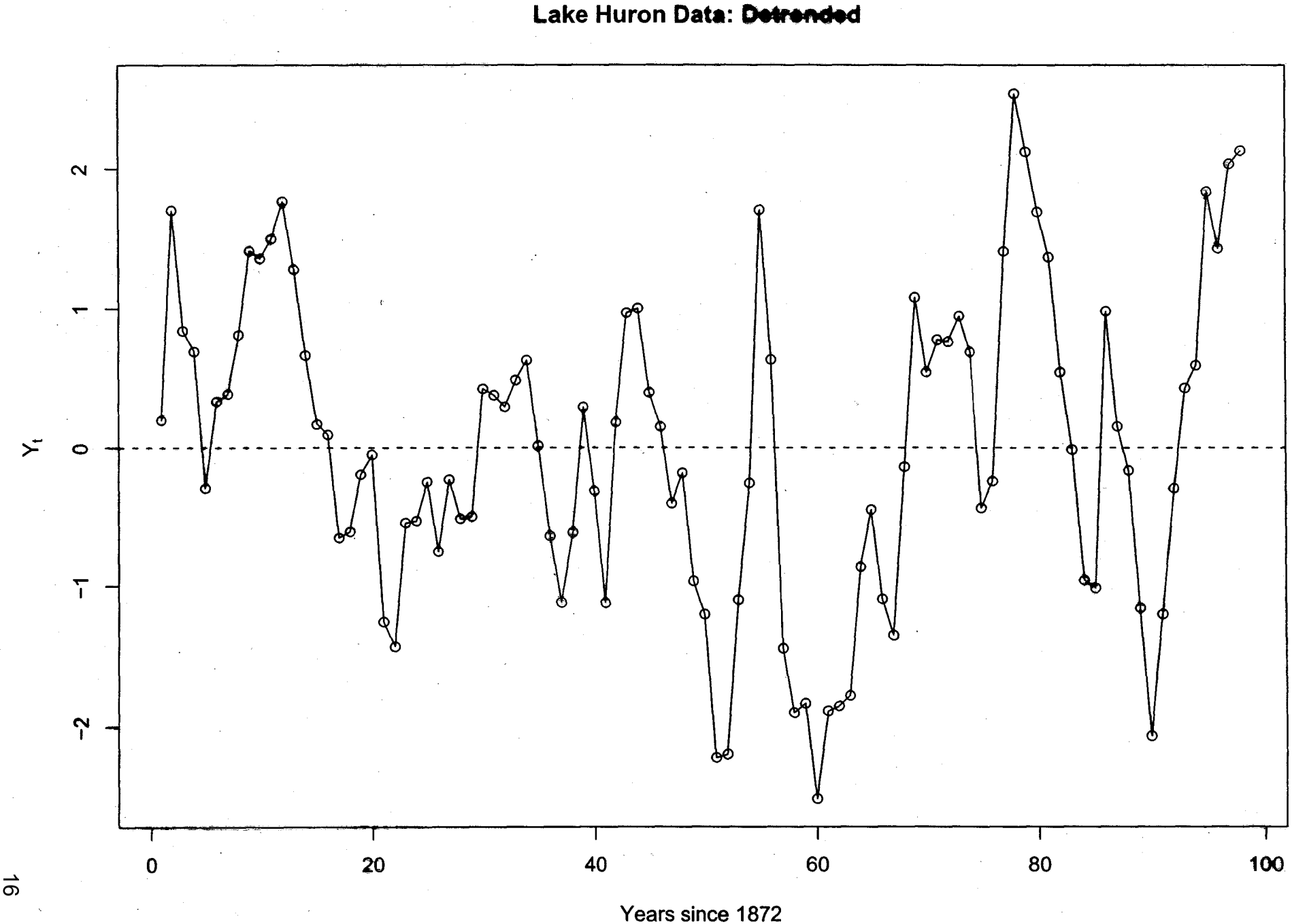
Resulting plot of $(t, \hat{y}_t)$ shows that resulting "residual" series $\{\hat{y}_t\}$ is a zero-mean process.

However, the residual series exhibits some structure

- successive values of

$$\hat{y}_1, \ldots, \hat{y}_T$$

have the same sign on many occasions!


Lake Huron Data: Detrended
Y_t
2
1
0
−1
−2
0
20
40
60
80
100
Years since 1872

A scatterplot of

$$(\hat{y}_{t-1}, \hat{y}_t)$$

pairs reveals apparent strong

positive correlation

Sample correlation of successive residuals

$$\hat{r} = \frac{1}{n-1} \sum_{t=2}^{98} (\hat{y}_t \hat{y}_{t-1} - \bar{y}_{(1)} \bar{y}_{(98)})$$

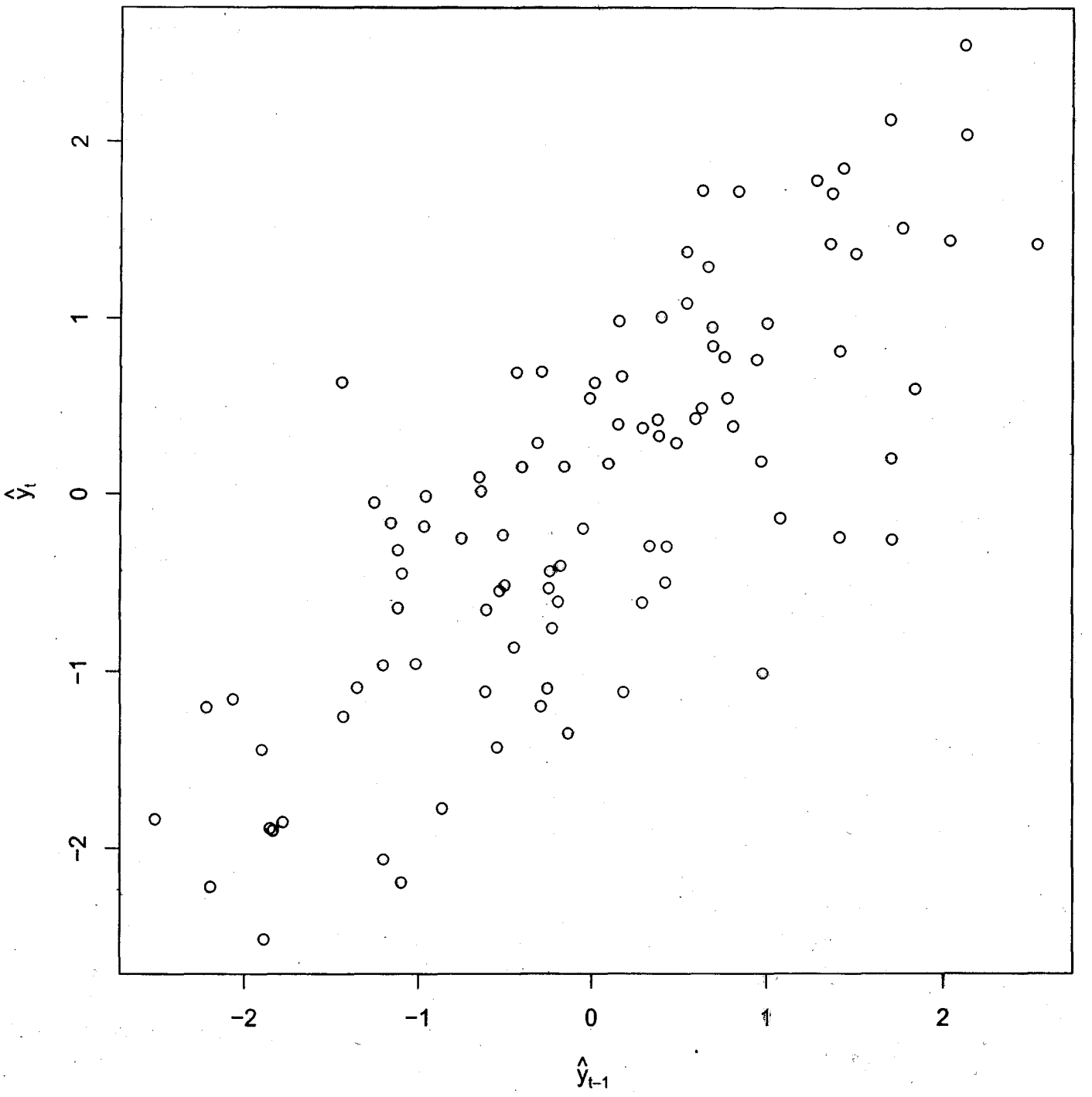
where $\bar{y}_{(1)} = \frac{1}{n-1} \sum_{t=2}^{98} \hat{y}_t$

$$\bar{y}_{(98)} = \frac{1}{n-1} \sum_{t=1}^{97} \hat{y}_t$$

(after dividing by sample variance)

$= 0.775.$

Lake Huron Data: Scatterplot of successive residuals



This positive dependence is useful in forecasting:

- estimated residuals at

$t = 93, 94, \ldots, 98$

are all positive

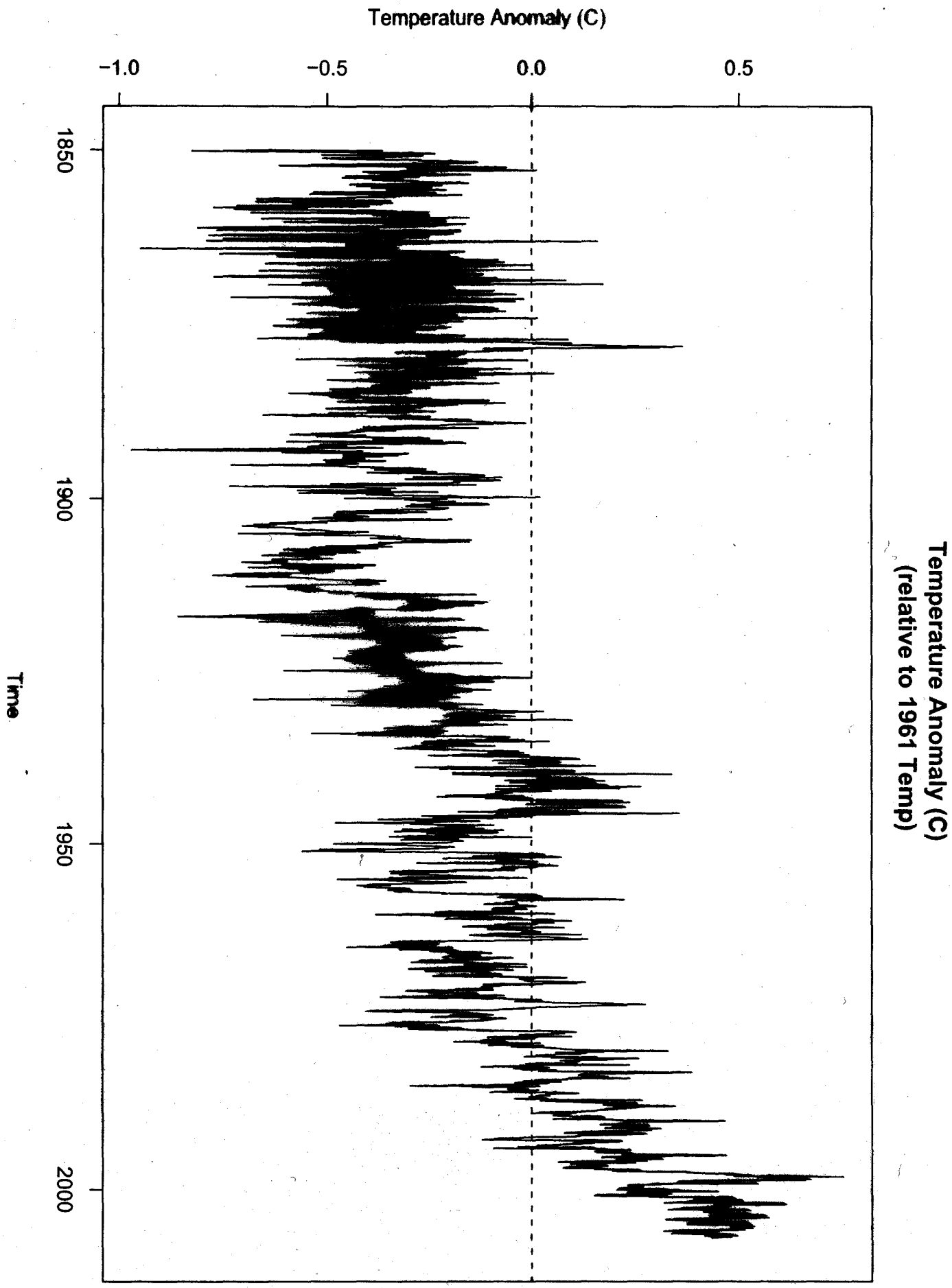
$\therefore$ it seems likely that

$\hat{Y}_{99}$

will also be positive

We seek models that capture this positive dependence

(or negative dependence if necessary)


Temperature Anomaly (C)
−1.0
−0.5
0.0
0.5
1850
1900
1950
2000
Time
Temperature Anomaly (C)
(relative to 1961 Temp)

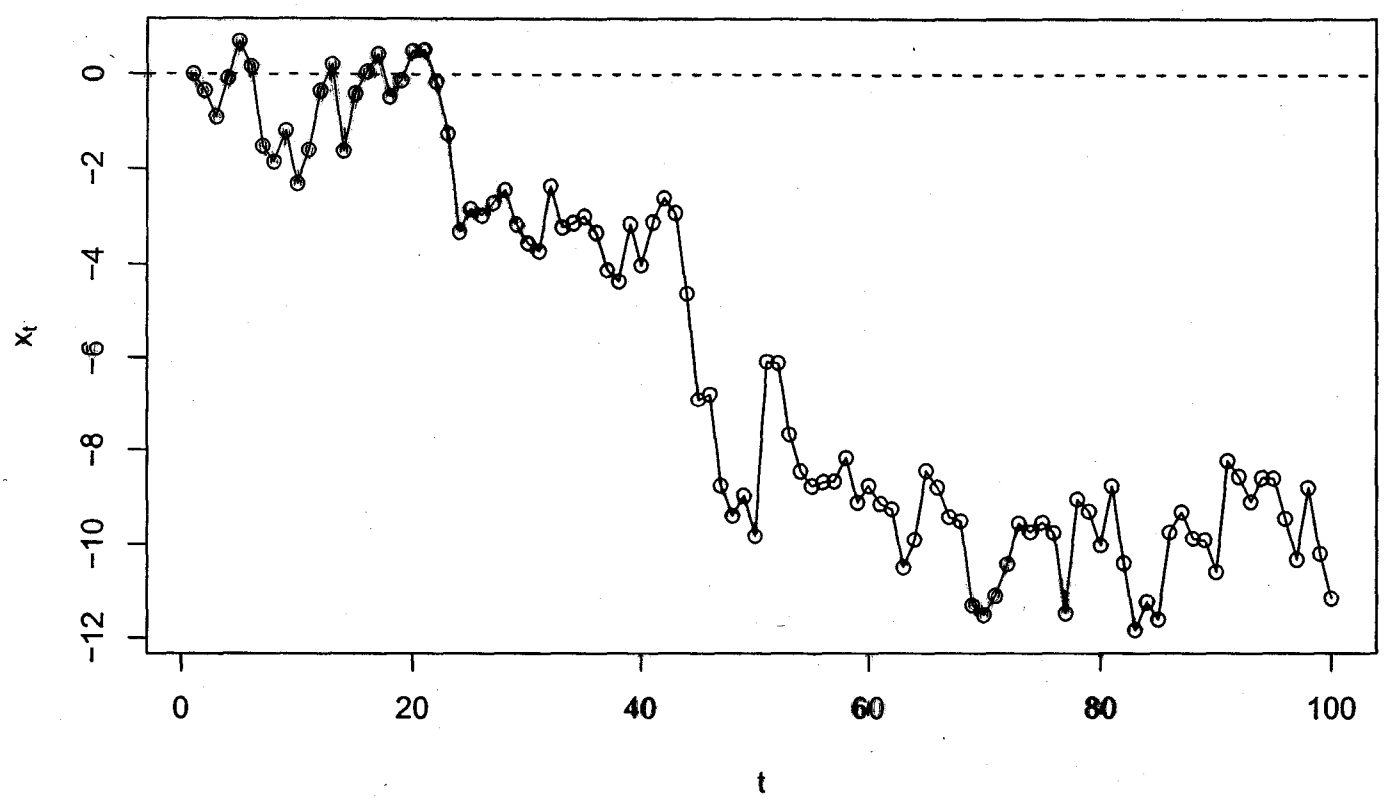
NOTE :

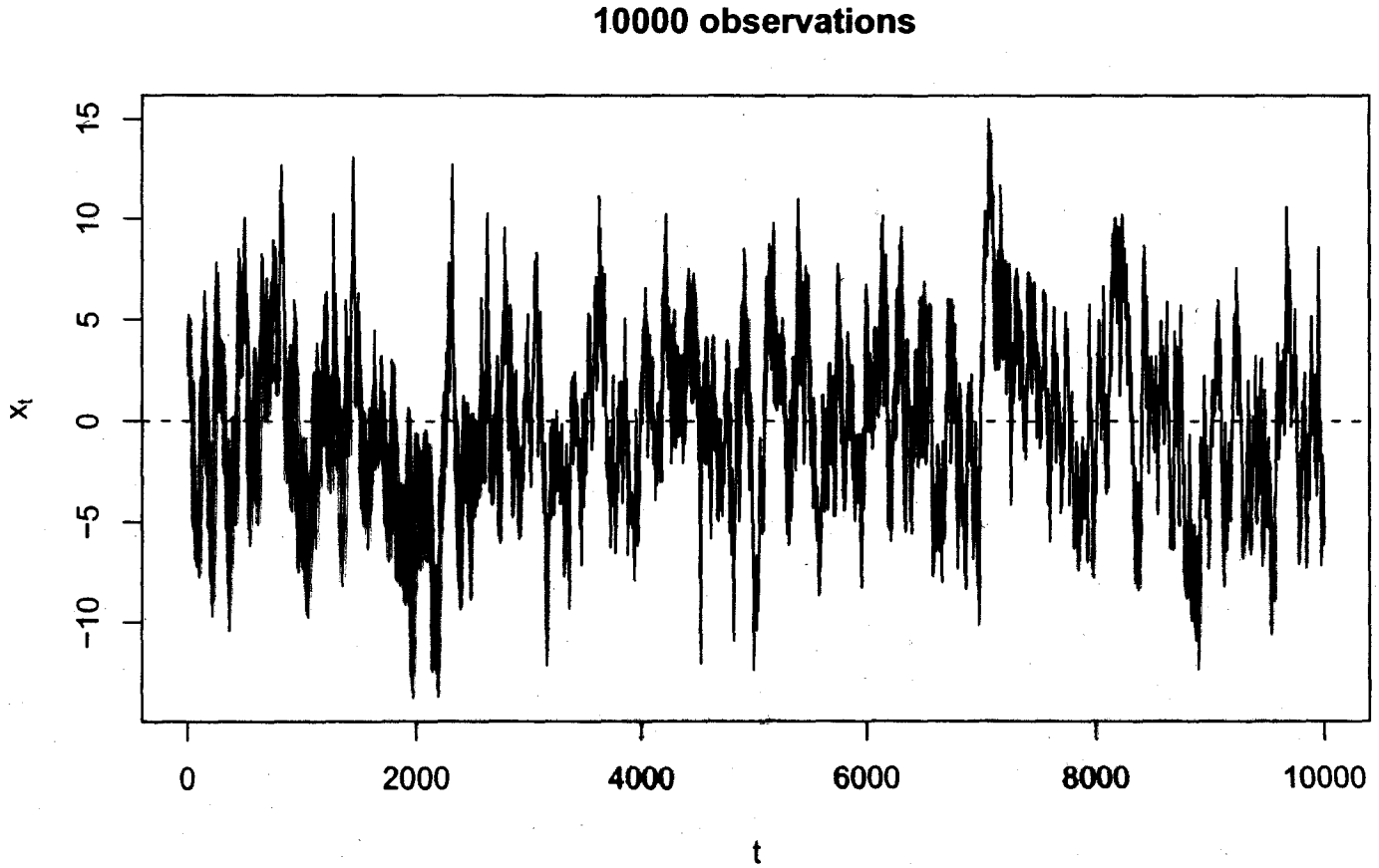
① Other models for trend may be fitted

- piecewise constant
- spline

② An apparent trend is not always a trend ....

.... and the decision to detrend may have important implications in terms of forecasting.


100 observations
x_t
0
-2
-4
-6
-8
-10
-12
0
20
40
60
80
100
t


10000 observations
x_t
15
10
5
0
-5
-10
0
2000
4000
6000
8000
10000
t

SEASONALITY

Many time series are influenced by "seasonally" varying factors

- calendar
- climate
- economic cycles

We can attempt to model seasonal variability using a deterministic (regression-type) model

i.e $$X_t = s_t + Y_t$$

where s_t is a periodic function with period d (i.e. $s_{t-d} = s_t \quad \forall t$)

In a "harmonic regression", s_t is modelled using a sum of sine waves ("harmonics")

$$s_t = a_0 + \sum_{j=1}^{k} \left[a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t) \right]$$

where

a_0

$a_1, \ldots, a_k$

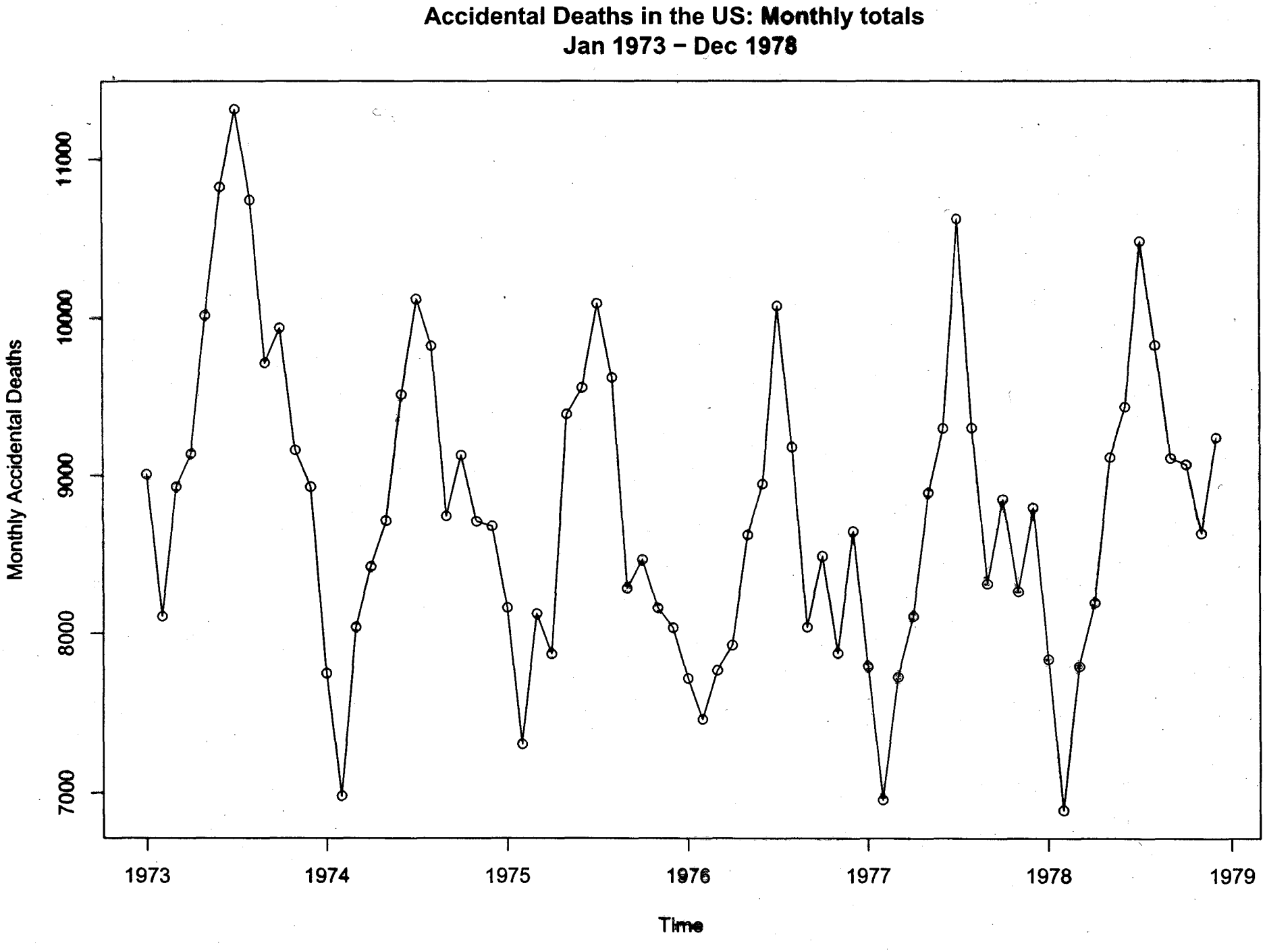
$b_1, \ldots, b_k$

are <u>unknown</u> coefficients

$\lambda_1, \ldots, \lambda_k$

are <u>fixed</u> frequencies (<u>known</u> constants)

[NOTE: STILL A "LINEAR MODEL"]


Accidental Deaths in the US: Monthly totals
Jan 1973 – Dec 1978
Monthly Accidental Deaths
7000
8000
9000
10000
11000
1973
1974
1975
1976
1977
1978
1979
Time

For Accidental Deaths data

$T = 72$, monthly data

To fit a model with

$k = 2$

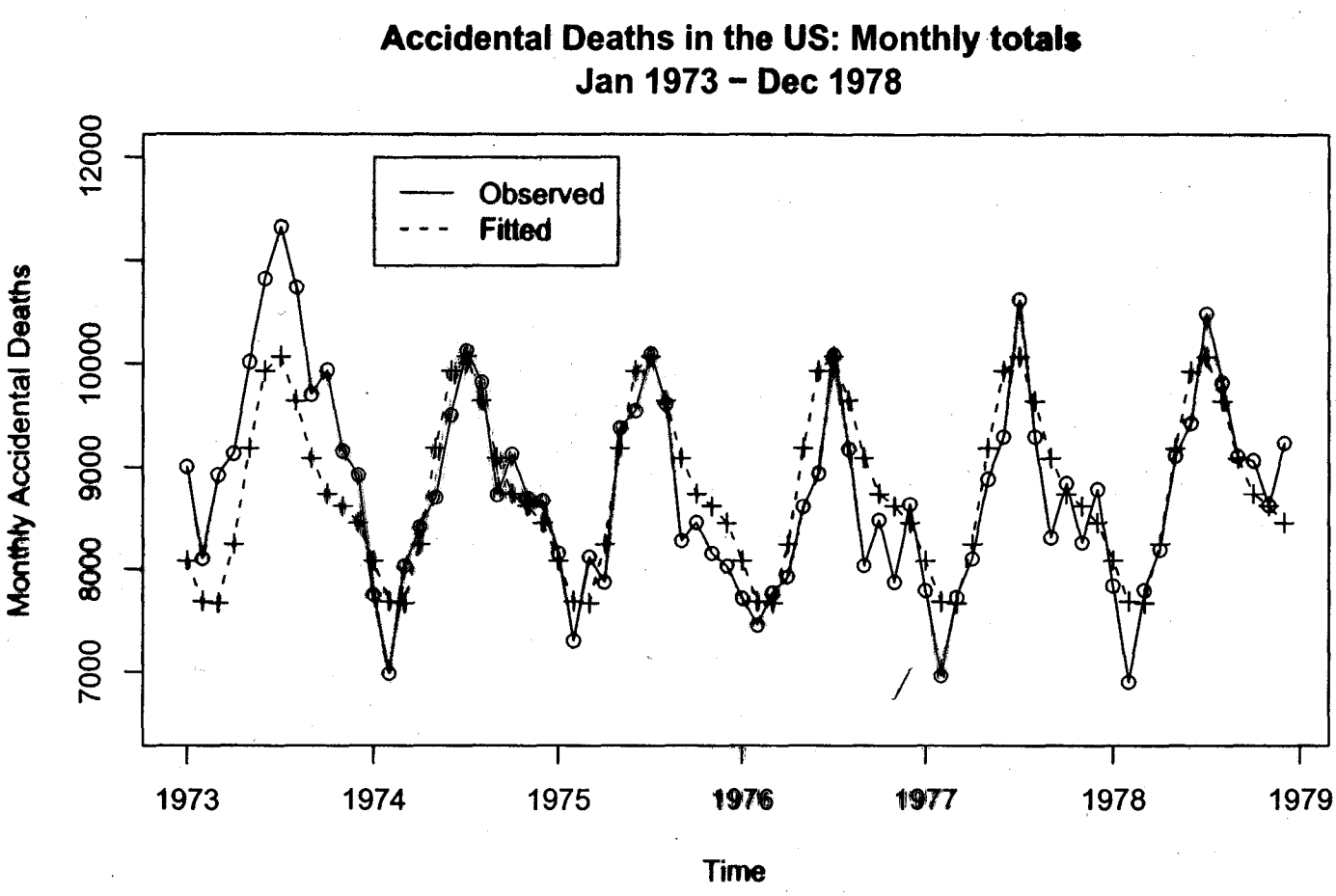
12 monthly cycle

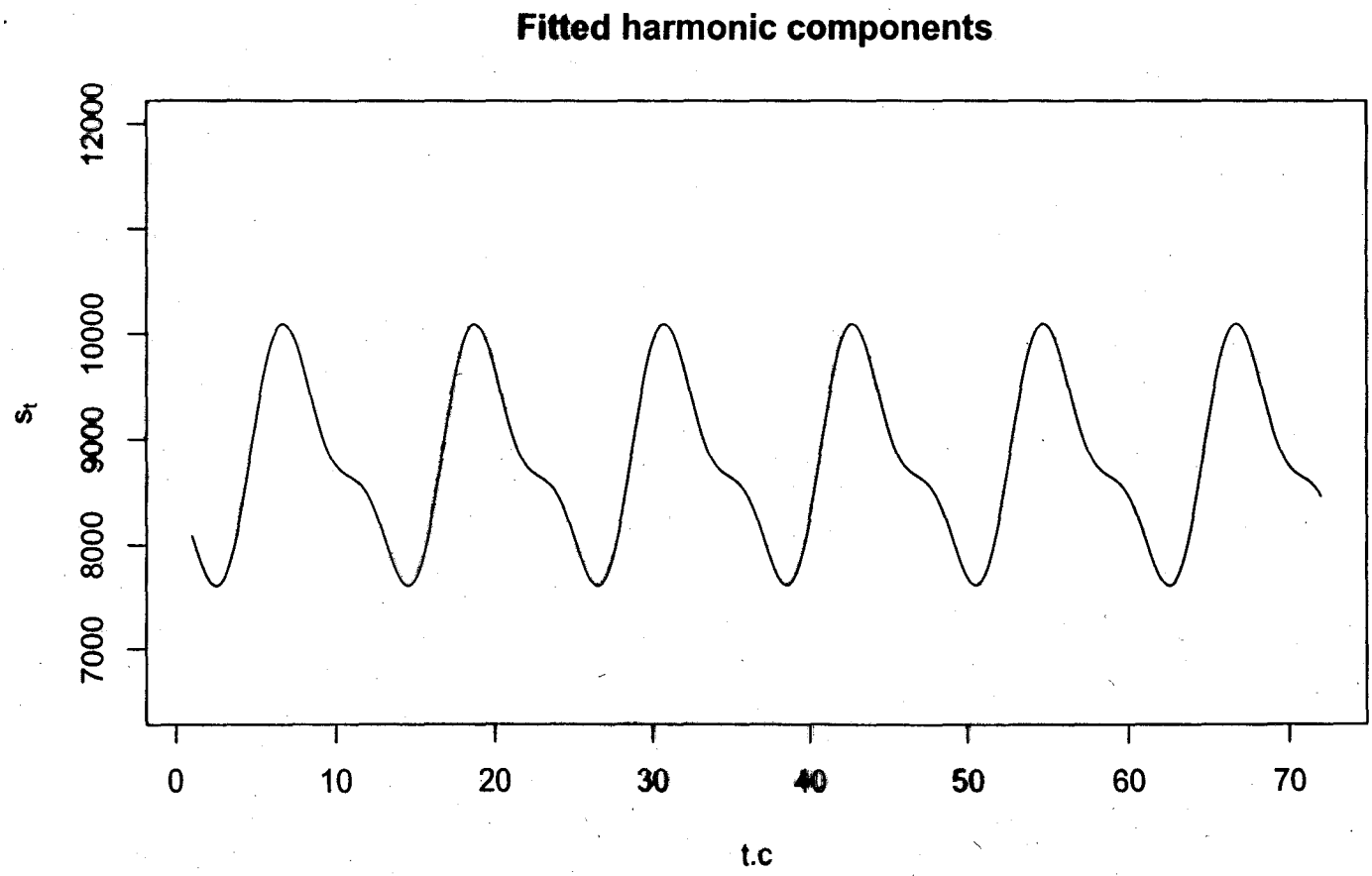
6 monthly cycle

set

$$\lambda_1 = \frac{2\pi}{12}$$

$$\lambda_2 = \frac{2\pi}{6}$$


Accidental Deaths in the US: Monthly totals
Jan 1973 – Dec 1978
Observed
Fitted
Monthly Accidental Deaths
7000
8000
9000
10000
12000
1973
1974
1975
1976
1977
1978
1979
Time


Fitted harmonic components
s_t
7000
8000
9000
10000
12000
0
10
20
30
40
50
60
70
t.c

We obtain

$\hat{a}_0 = 8788.79 \quad (66.25)$

$\hat{a}_1 = -734.19 \quad (93.69)$ } 12 MONTH COMPON.

$\hat{b}_1 = -711.32 \quad (93.69)$

$\hat{a}_2 = 408.04 \quad (93.69)$ } 6 MONTH COMP.

$\hat{b}_2 = 97.12 \quad (93.89)$

b_2 not significantly different from zero

Fit of model is adequate except near beginning of the series

"Residuals" again appear to exhibit positive correlation

Stationarity

A time series is <u>stationary</u> if

$$\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$$

and

$$\{X_{t+h}, t = 0 \pm 1, \pm 2, \ldots\}$$

for integer values of h have the same statistical properties.

- joint dist^n
- second order properties

DEFINITION Let $\{X_t\}$ be a timeseries with $E[X_t^2] < \infty$. Then define

- the **mean-function** $\mu_X(t) = E[X_t]$

- the **covariance function**

$$\gamma_X(r,s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

$$= \text{Cov}[X_r, X_s]$$

for integers r, s.

DEFINITION $\{X_t\}$ is **weakly stationary** if

(i) $\mu_X(t)$ is independent of t

(ii) $\gamma_X(t+h, t)$ is independent of t for each h.

DEFINITION $\{X_t\}$ is **strictly stationary** if

$$(X_{t+1}, \ldots, X_{t+n})$$

and

$$(X_{t+h+1}, \ldots, X_{t+h+n})$$

have the same joint distribution for all t, h, n.

DEFINITION

For weakly stationary processes, we denote the autocovariance function (ACVF)

$$\gamma_X(h) \equiv \gamma_X(h,0) = \gamma_X(t+h,t)$$

and the autocorrelation function (ACF)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

$$\gamma_X(h) = Cov[X_{t+h}, X_t]$$

$$\rho_X(h) = Corr[X_{t+h}, X_t]$$

Bilinearity property

$$\mathrm{Cov}[aX+bY+c, Z]$$

$$= a\,\mathrm{Cov}[X,Z] + b\,\mathrm{Cov}[Y,Z]$$

for a, b, c constants

$$E[X^2], E[Y^2], E[Z^2] < \infty.$$

This will be the central tool for computing

$$\gamma_X(h), \rho_X(h).$$

EXAMPLE IID NOISE $\{X_t\} \sim IID(0, \sigma^2)$

IID noise process is characterized by

$$E[X_t] = 0$$

$$E[X_t^2] = \sigma^2 < \infty$$

with all X_t <u>independent</u>

$$\gamma_X(h) = \gamma_X(t+h, t) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

EXAMPLE WHITE NOISE $\{X_t\} \sim WN(0, \sigma^2)$

$$E[X_t] = 0$$
$$E[X_t^2] = \sigma^2$$

with X_t uncorrelated

Both IID NOISE and WHITE NOISE processes are stationary.

EXAMPLE RANDOM WALK

$$\{X_t\} \sim IID(0, \sigma^2)$$

$\{S_t\}$ defined by $S_t = \sum_{i=1}^{t} X_i$.

Then $E[X_t] = 0 \Rightarrow E[S_t] = 0$

$E[X_t^2] = \sigma^2 \Rightarrow E[S_t^2] = t\sigma^2$

$\therefore$ S_t is non-stationary

$$\gamma_S(t+h, t) = \mathrm{Cov}[S_{t+h}, S_t]$$

$$= \mathrm{Cov}[S_t + X_{t+1} + \ldots + X_{t+h}, S_t]$$

$$= \mathrm{Cov}[S_t, S_t]$$

$$+ \sum_{i=1}^{h} \underbrace{\mathrm{Cov}[X_{t+i}, S_t]}_{=0 \text{ as } \{X_t\} \sim IID}$$

$$= \mathrm{Cov}[S_t, S_t]$$

$$= t\sigma^2$$

- depends on t.

EXAMPLE MOVING AVERAGE $\{X_t\} \sim$ [illegible]

Suppose $\{Z_t\}$ is a $WN(0, \sigma^2)$ process

$\theta \in \mathbb{R}$.

$\sigma^2 < \infty$.

Let

$$X_t = Z_t + \theta Z_{t-1} \qquad t = 0, \pm 1, \pm 2, \ldots$$

Then

$$E[X_t] = 0$$

and

$$\gamma_X(t+h, t) = Cov[X_{t+h}, X_t]$$

$$= E[X_{t+h} X_t]$$

Now

$$X_{t+h} X_t = \left(Z_{t+h} + \theta Z_{t+h-1}\right)\left(Z_t + \theta Z_{t-1}\right)$$

$$= Z_{t+h} Z_t + \theta\left(Z_{t+h} Z_{t-1} + Z_{t+h-1} Z_t\right) + \theta^2 Z_{t+h-1} Z_{t-1}$$

$$E[Z_{t+h} Z_t] = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

$$E[Z_{t+h} Z_{t-1}] = \begin{cases} \sigma^2 & h = -1 \\ 0 & h \neq -1 \end{cases}$$

$$E[Z_{t+h-1} Z_t] = \begin{cases} \sigma^2 & h = 1 \\ 0 & h \neq 1 \end{cases}$$

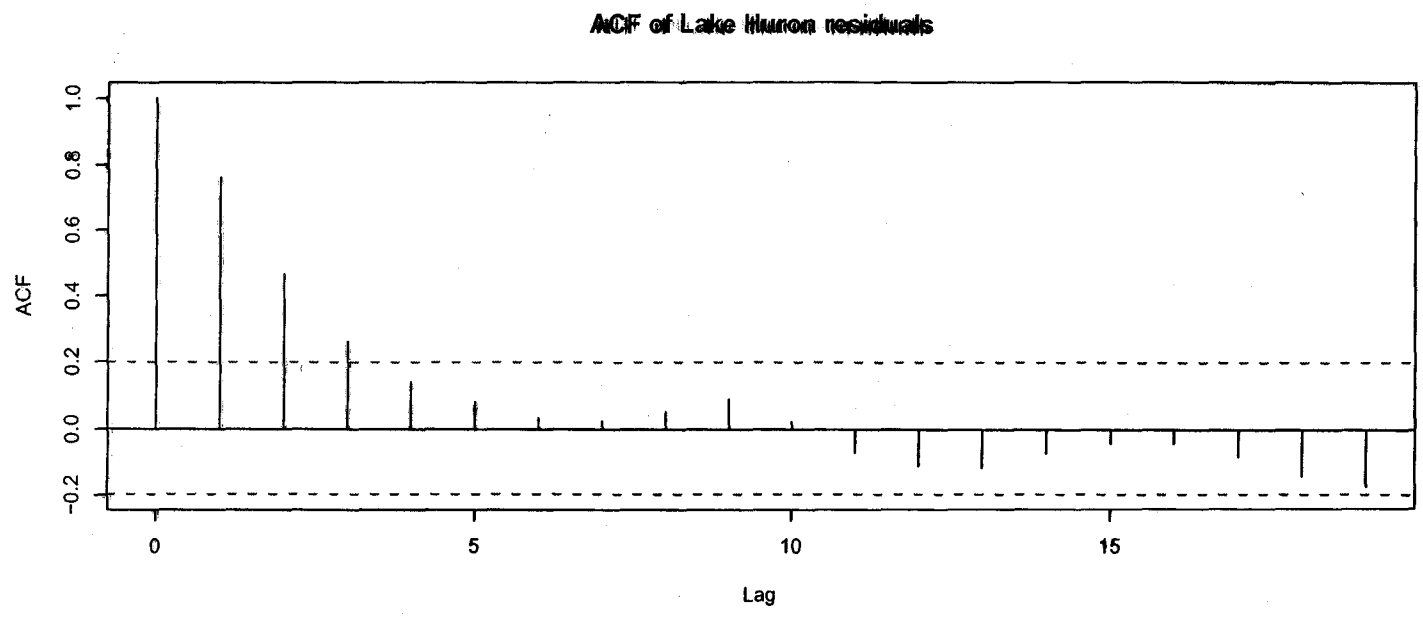
$$E[Z_{t+h-1} Z_{t-1}] = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

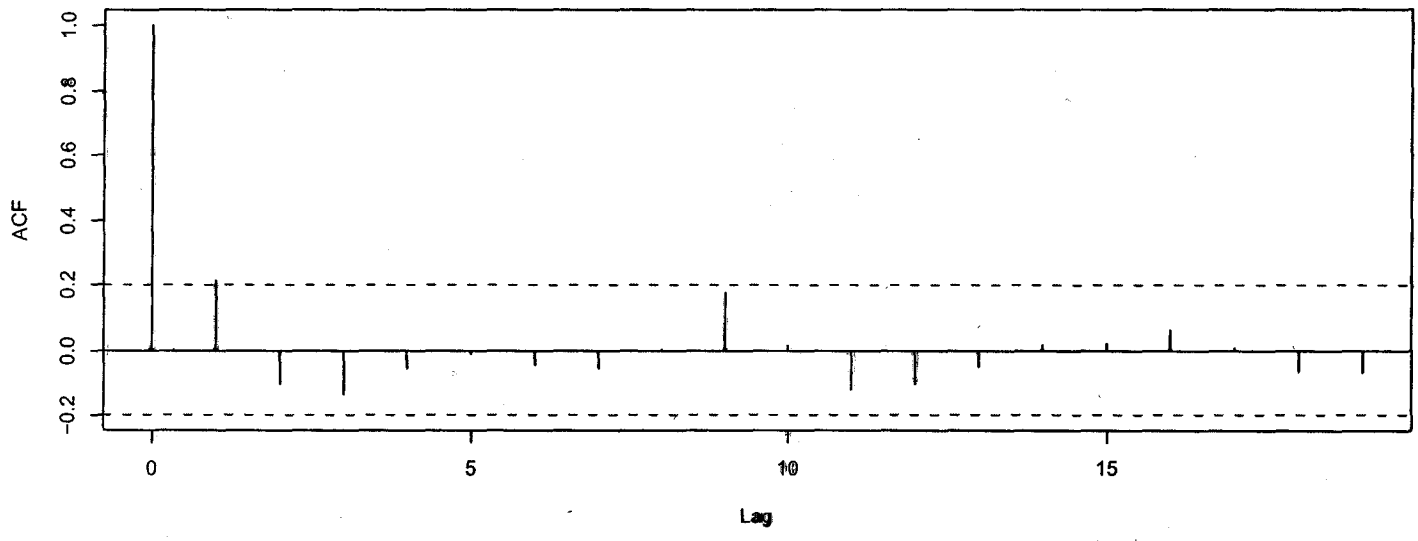
$$\therefore \quad \gamma_X(t+h, t) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \sigma^2\theta & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

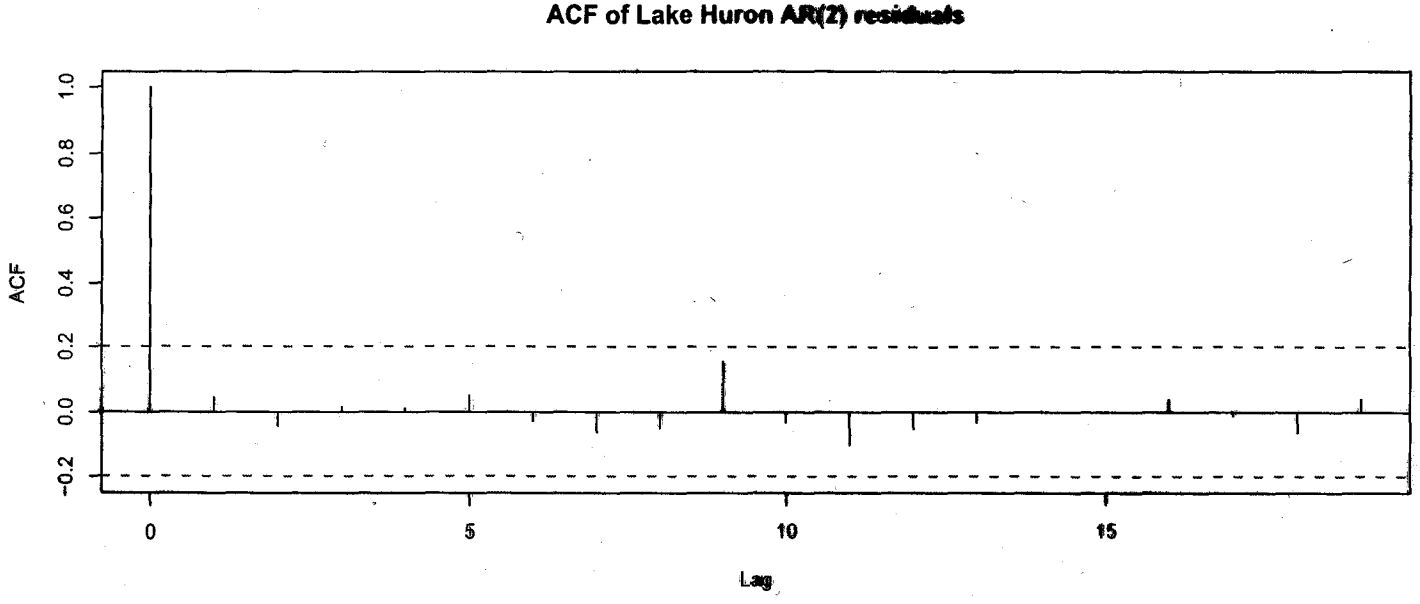
$$\therefore \quad \rho_X(t+h, t) = \begin{cases} 1 & h = 0 \\ \dfrac{\theta}{1+\theta^2} & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$ $\{X_t\}$ is a stationary process

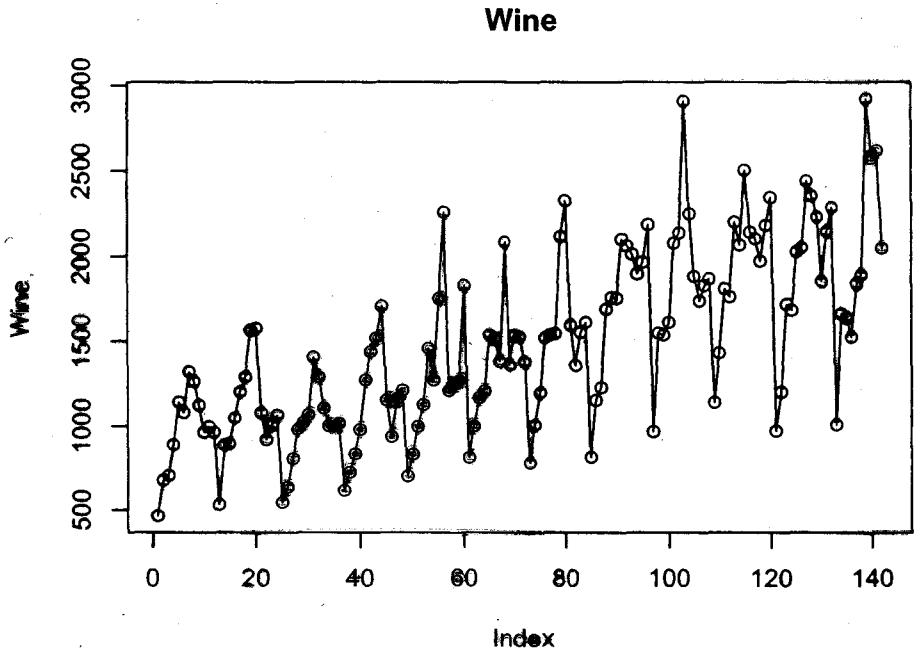
$$\{X_t\} \sim MA(1)$$

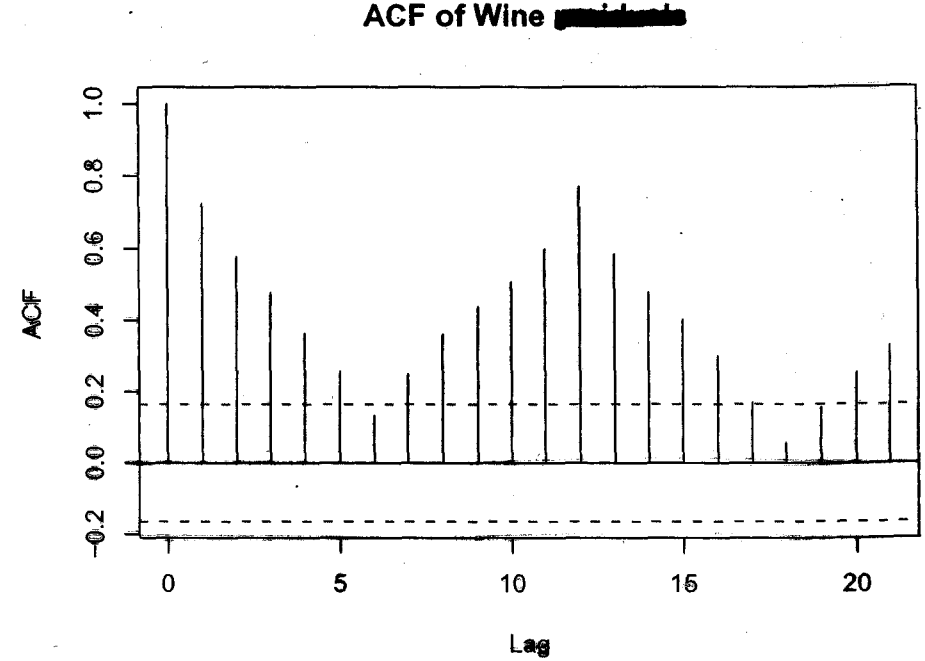

ACF of Lake Huron residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

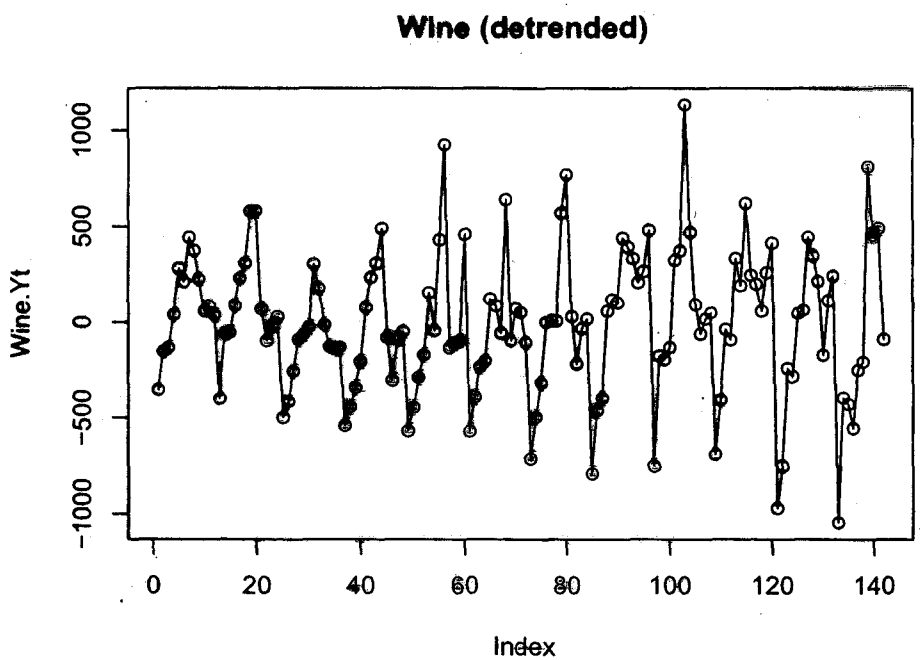

ACF of Lake Huron AR(1) residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

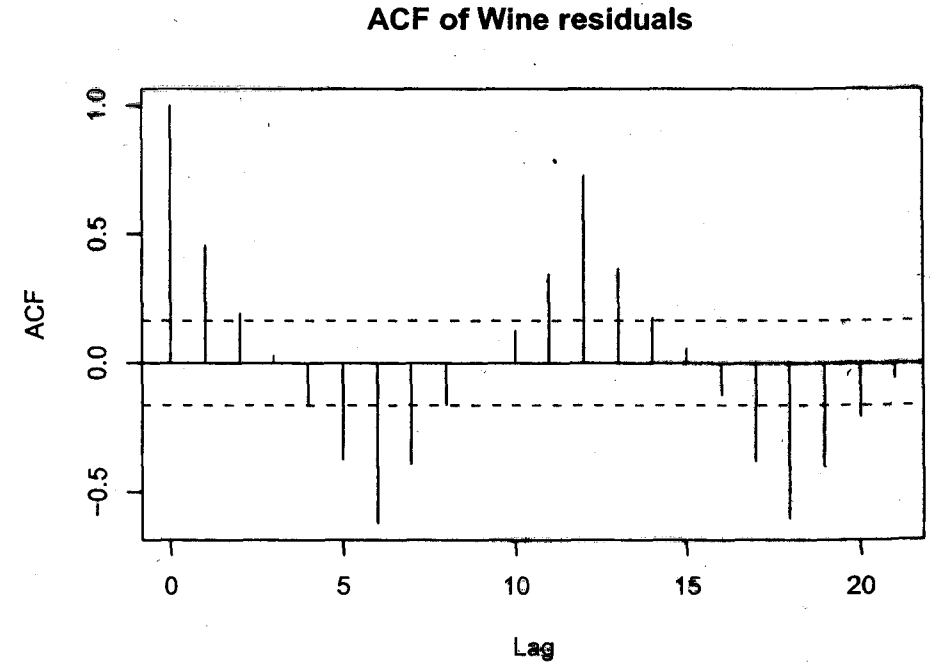

ACF of Lake Huron AR(2) residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

0.9
0.2
0.1
Strikes
3500
4000
4500
5000
5500
6000
t.Strikes
0
5
10
15
20
25
30


Wine
Wine
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140
Index


ACF of Wine
ACF
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Lag


Wine (detrended)
Wine.Yt
-1000
-500
0
500
1000
0
20
40
60
80
100
120
140
Index


ACF of Wine residuals
ACF
-0.5
0.0
0.5
1.0
0
5
10
15
20
Lag

Strikes
3500
4000
4500
5000
5500
6000
t.Strikes
0
5
10
15
20
25
30

EXAMPLE AUTOREGRESSION $\{X_t\} \sim AR(p)$

Suppose $\{Z_t\} \sim WN(0, \sigma^2)$

$|\phi| < 1$

and that $\{X_t\}$ is stationary such that Z_t and X_s are uncorrelated if $s < t$, with

$$X_t = \phi X_{t-1} + Z_t \qquad t = 0, \pm 1, \pm 2, \ldots$$

Then

$$E[X_t] = \phi E[X_{t-1}] + E[Z_t]$$

$$= \phi E[X_{t-1}]$$

$\therefore$ $E[X_t] = 0$ as $\{X_t\}$ is stationary

$$\gamma_x(h) = Cov[X_{t+h}, X_t]$$

$$= Cov[X_t, X_{t-h}]$$

$$= Cov[\phi X_{t-1}, X_{t-h}] + Cov[Z_t, X_{t-h}]$$

$$= \phi\, Cov[X_{t-1}, X_{t-h}] + 0$$

$$= \phi\, \gamma_x(h-1)$$

$$= \phi^2 \gamma_x(h-2)$$

$$= \ldots$$

$$= \phi^h \gamma_x(0)$$

As $\gamma_x(h) = \gamma_x(-h)$, we have

$$\rho_x(h) = \phi^{|h|} \qquad h = 0, \pm 1, \pm 2, \ldots$$

Note also that

$$\gamma_X(0) = \text{Cov}[X_t, X_t]$$

$$= \text{Cov}[\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t]$$

$$= \phi^2 \text{Cov}[X_{t-1}, X_{t-1}] + \text{Cov}[Z_t, Z_t]$$

$$= \phi^2 \gamma_X(0) + \sigma^2$$

$$\therefore \gamma_X(0) = \frac{\sigma^2}{1-\phi^2}$$

We may use moment-based estimates of these quantities for stationary processes:

$\bar{x}$ to estimate μ

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

$$-n < h < n$$

to estimate $\gamma_X(h)$

- consistent but not unbiased
- ensures sample covariance matrix

$$\hat{\Gamma}_n = [\hat{\gamma}(i-j)]_{ij}$$

is non-negative definite.

EXAMPLE LAKE HURON DATA

$$X_t = m_t + Y_t$$

- estimate m_t using linear trend.
- left residuals that were positively correlated
- $\text{Corr}[\hat{y}_{t-1}, \hat{y}_t] \approx 0.77$

Suggests an AUTOREGRESSION model for $\{Y_t\}$

Fit a linear regression model to $\{\hat{y}_t\}$ of the form

$$\hat{y}_t = \phi \hat{y}_{t-1} + z_t$$

- fit using OLS
- yields

$$\hat{\phi} = 0.791$$

- sample variance of

$$\hat{z}_t = \hat{y}_t - \hat{\phi}\hat{y}_{t-1}$$

is 0.5024

Is $\hat{z}_t$ an uncorrelated sequence?

- apparently not!
- still positive correlation between $\hat{z}_{t-1}, \hat{z}_t$.

$\therefore \quad Y_t = \phi Y_{t-1} + z_t \qquad \{z_t\} \sim IID$

does <u>not</u> fit the observed data.

A better model is a

SECOND-ORDER AUTOREGRESSION

AR(2)

Consider the model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$$

where $\{Z_t\}$ is IID.

Again, this model can be fit using OLS to the $\{\hat{y}_t\}$ sequence.

- yields

$$\hat{\phi}_1 = 1.002$$

$$\hat{\phi}_2 = -0.2834$$

The estimated variance of $\{\hat{z}_t\}$ is

0.4460

and the estimated residual sequence is approximately uncorrelated.

Thus we have a range of techniques available for

- trend
- seasonality

removal. In the general model

$$X_t = m_t + s_t + Y_t$$

we can model m_t and s_t, leaving $\{Y_t\}$ a zero-mean stationary process which is

- IID
- WN
- MA
- AR.

TREND REMOVAL :

- LINEAR MODEL + OLS to estimate trend.

- "SMOOTHING" using MOVING AVERAGE

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t-q}$$

for $q+1 \le t \le T-q$

(with $X_s \equiv X_1$ for $s \le 0$

$X_s \equiv X_T$ for $s > T$)

This form of smoothing involves the "low pass" filter

$$\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$$

$$a_j = \begin{cases} \frac{1}{2q+1} & -q \le j \le q \\ 0 & \text{otherwise} \end{cases}$$

- different filters appropriate for different sorts of trend.

- "EXPONENTIAL" SMOOTHING

for some α $\quad 0 < \alpha < 1$

$$\hat{m}_t = \begin{cases} \alpha X_t + (1-\alpha)\hat{m}_{t-1} & t \geq 2 \\ X_1 & t = 1 \end{cases}$$

$\therefore$

$$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha) X_{t-j} + (1-\alpha)^{t-1} X_1$$

by recursion.

α small - low smoothing

α large - extreme smoothing

GENERAL STRATEGY "CLASSICAL"

Suppose

$$X_t = m_t + s_t + Y_t$$

with $E[Y_t] = 0$

$$s_{t+d} = s_t$$ SEASONALITY PERIOD d

$$\sum_{j=1}^{d} s_j = 0$$

STEP 1 If

d even	set	$q = \frac{d}{2}$
d odd	set	$q = \frac{d-1}{2}$

d even

$$\hat{m}_t = \frac{x_{t-q} + x_{t+q}}{2d} + \frac{1}{d} \sum_{j=-(q-1)}^{q-1} x_{t+j}$$

d odd

$$\hat{m}_t = \frac{1}{d} \sum_{j=-q}^{q} x_{t+j}$$

- REMOVES TREND RESPECTING SEASONALITY

STEP 2 Remove the seasonality using

$$w_k = \frac{1}{n_j} \sum_j \left(x_{k+jd} - \hat{m}_{k+jd} \right)$$

where the sum extends over

$$q < k+jd \leq n-q$$

and n_j is the number of terms in the sum. Then

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{i=1}^{d} w_i \qquad k=1,\ldots,d.$$

and

$$\hat{s}_k = \hat{s}_{k-d} \qquad k > d.$$

The deseasonalized data are

$$d_t = x_t - \hat{s}_t \qquad t = 1, \ldots, n$$

Step 3 Estimate the "noise" process

$$\hat{y}_t = x_t - \hat{m}_t^* - \hat{s}_t$$

where the trend estimate $\hat{m}_t^*$ is estimated from the deseasonalized data $\{d_t\}$.

DIFFERENCING

Differencing is another method that can be used to transform a non-stationary trending time series into a stationary one.

DEFINITION DIFFERENCE OPERATOR

The lag-1 difference operator, ∇, by

$$\nabla X_t = X_t - X_{t-1}$$

$$= (1 - B)X_t$$

where B is the backshift operator

$$BX_t = X_{t-1}$$

Repeated application of these operators is denoted in the obvious way

$$B^j X_t = X_{t-j}$$

$$= B(X_{t-j+1}) \quad \text{etc}$$

$$\nabla^j X_t = \nabla(\nabla^{j-1}(X_t)) \quad \text{etc}$$

These operators are manipulated as polynomial functions

eg $\nabla^2 X_t = \nabla(\nabla X_t)$

$$= (1-B)(1-B)X_t$$

$$= (1-2B+B^2)X_t$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

In general, differencing can be used to remove polynomial trends

eg $m_t = a_0 + a_1 t$

$\Rightarrow \nabla m_t = m_t - m_{t-1}$

$= (a_0 + a_1 t) - (a_0 + a_1 (t-1))$

$= a_1$

To remove a polynomial trend, order k, apply k^{th} order differencing

∇^k

SEASONAL DIFFERENCING

DEFINITION LAG-d DIFFERENCE OPERATOR

The lag-d difference operator, ∇_d, is defined by

$$\nabla_d X_t = X_t - X_{t-d}$$

$$= (1 - B^d) X_t$$

This operator can remove a seasonality with period d: if s_t has period d and

$$X_t = m_t + s_t + Y_t$$

then $\nabla_d X_t = (m_t - m_{t-d}) + \underbrace{(s_t - s_{t-d})}_{=0} + (Y_t - Y_{t-d})$

$$= (m_t - m_{t-d}) + (Y_t - Y_{t-d})$$

Can combine ∇ and ∇_d

$$\nabla \nabla_d X_t = (1-B)(1-B^{12}) X_t$$

$$= (1 - B - B^{12} + B^{13}) X_t$$

$$= X_t - X_{t-1} - X_{t-12} + X_{t-13}$$

- removes seasonal and trend components
- note that

$$\nabla \nabla_d X_t = \nabla_d \nabla X_t$$

The objective of the "transformations" is to focus on the zero-mean process $\{Y_t\}$ in

$$X_t = m_t + s_t + Y_t$$

with $\{Y_t\}$ stationary. A further decomposition represents $\{Y_t\}$ using an IID/WN/MA/AR form:

$$Y_t = g(\tilde{Y}_{t-1}, \tilde{Z}_{t-1}) + Z_t$$

where $\{Z_t\}$ is a $WN(0, \sigma^2)$ process

$$\tilde{Y}_t = (Y_1, \ldots, Y_t)$$

$$\tilde{Z}_t = (Z_1, \ldots, Z_t)$$

g is a (linear) function

TESTING THE SEQUENCE $\{Z_t\}$

Four established testing procedures for assessing whether $\{Z_t\}$ is $IID(0,\sigma^2)$

$$H_0 : \{Z_t\} \sim IID(0,\sigma^2)$$
$$H_1 : \{Z_t\} \not\sim IID(0,\sigma^2)$$

1. SAMPLE ACF TEST

If $\{Z_t\} \sim IID$, then as $n \to \infty$

$$\hat{\rho}_n(h) \dot{\sim} N\left(0, \frac{1}{n}\right)$$

where $\hat{\rho}_n(h)$ is the ACF estimated from data $z_1, \ldots, z_n$

i.e. each $\hat{\rho}_n(h)$ should lie in a range

$$\pm \frac{1.96}{\sqrt{n}}$$

- can base the test singly at one h value or simultaneously across h values ....

2. THE PORTMANTEAU TEST

Test statistic: $Q = n \sum_{j=1}^{h} [\hat{\rho}_n(h)]^2$

Under H_0, for large n

$$Q \dot{\sim} \chi^2_h$$

Ljung-Box Statistic:

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\{\hat{\rho}_n(j)\}^2/(n-j)$$

McLeod-Li Statistic

Let $w_t = z_t^2$ and consider the sample acf formed from the w_t values, denoted

$$\hat{\rho}_n^w(h)$$

Define

$$Q_{ML} = n(n+2)\sum_{j=1}^{h}\{\hat{\rho}_n^w(j)\}^2/(n-j)$$

Under H_0, for large n,

$$Q_{LB}, Q_{ML} \dot{\sim} \chi_h^2.$$

3. TURNING POINT TEST

i is a turning point if

$$z_{i-1} < z_i \quad \text{and} \quad z_i > z_{i+1}$$
$$\text{or} \quad z_{i-1} > z_i \quad \text{and} \quad z_i < z_{i+1}$$

If T_n is the number of turning points

$$T_n \dot{\sim} \text{Normal}(\mu_{T_n}, \sigma^2_{T_n})$$

$$\mu_{T_n} = \frac{2}{3}(n-2)$$

$$\sigma^2_{T_n} = \frac{16n - 29}{90}$$

$T_n < \mu_n \Rightarrow$ positive correlation

$T_n > \mu_n \Rightarrow$ negative correlation

4. DIFFERENCE - SIGN TEST

$$S_n = \sum_{i=2}^{n} I_{[0,\infty)}(z_i - z_{i-1})$$

Under the hypothesis of zero correlation

$$S_n \doteq \text{Normal}(\mu_{S_n}, \sigma_{S_n}^2)$$

$$\mu_{S_n} = \frac{1}{2}(n-1)$$

$$\sigma_{S_n}^2 = \frac{n+1}{12}$$

Tests also exist for

- trend
- normality

If a parametric model is fitted

eg $$X_t = \phi X_{t-1} + Y_t \qquad \{Y_t\} \sim IID(0, \sigma^2)$$

Then tests for ϕ may be of interest

$$H_0 : \phi = 0$$
$$H_1 : \phi \neq 0$$

$$H_0 : \phi = 1 \qquad \text{NON STATIONARITY}$$
$$H_1 : \phi < 1 \qquad \text{STATIONARITY}$$

2. Stationary Processes

We will focus on weakly stationary processes which are characterized by their mean and ACVS/ACF.

For a stationary process $\{X_t\}$

- $E[X_t] = \mu$, $Var[X_t] = \sigma^2$
- $\gamma_X(h)$; $\rho_X(h)$ (ACVS, ACF)

Consider prediction of

$$X_{n+h}$$

given that $X_n = x_n$.

If $\{X_t\}$ is a GAUSSIAN PROCESS

where

$$(X_{t+1}, \ldots, X_{t+s})$$

has a Multivariate Normal dist^n $\forall t, s$

then

$$X_{n+h} | X_n = x_n \sim N(m_h(x_n), v_h(x_n))$$

where

$$m_h(x_n) = \mu + \rho_x(h)(x_n - \mu)$$

$$v_h(x_n) = \sigma^2(1 - (\rho_x(h))^2)$$

We wish to predict X_{n+h} given $X_n = x_n$.

The "best prediction" criterion is MEAN-SQUARE ERROR...

Choose $\hat{X}_{n+h}$ to <u>minimize</u>

$$MSE = E\left[(X_{n+h} - \hat{X}_{n+h})^2 \mid X_n = x_n\right]$$

It follows that

$$\hat{X}_{n+h} = E\left[X_{n+h} \mid X_n = x_n\right]$$

$$= m_h(x_n) = \mu + \rho_X(h)(x_n - \mu)$$

- corresponding <u>predictor</u> is

$$m_h(X_n) = \mu + \rho_X(h)(X_n - \mu)$$

- corresponding MSE is

$$\sigma^2\left(1 - (\rho_X(h))^2\right)$$

For a non-Gaussian process we seek the best <u>linear</u> predictor

$$L_n(X_n) = aX_n + b$$

Again under MSE, the best predictor is

$$L_n(X_n) = \mu + \rho_X(h)(X_n - \mu)$$

with MSE

$$\sigma^2\left(1-(\rho_X(h))^2\right)$$

Thus in both cases $\rho_X(h)$ plays a crucial role.

A real-valued function g

$$g: \mathbb{Z} \to \mathbb{R}$$

is <u>non-negative definite</u> if $\forall$ integers n

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \, g(i-j) \, a_j \geq 0$$

for all $\underset{\sim}{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$

THEOREM

Function g defined on $\mathbb{Z}$ is the autocovariance function of a stationary time series if and only if it is <u>even</u> and <u>non-negative definite</u>.

Note: by properties of covariance

$$\gamma(0) \geq 0$$

$$|\gamma(h)| \leq \gamma(0)$$

$$\gamma(h) = \gamma(-h)$$

Proof

If: Let $\gamma(\,)$ be the acv function of a stationary time series.

Let $\underset{\sim}{a} = (a_1, \ldots, a_n)^T$

$\underset{\sim}{X}_n = (X_n, \ldots, X_1)^T$

Then

$$\mathrm{Var}\left[a^T \underset{\sim}{X}_n\right] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \gamma(i-j) a_j$$

$$= \underset{\sim}{a}^T \Gamma_n \underset{\sim}{a} \geq 0$$

where $[\Gamma_n]_{ij} = \gamma(i-j)$.

Only if: Given $\gamma(\,)$ with

$\gamma(h) = \gamma(-h)$

and $\gamma(\,)$ non-negative definite

we need to find stationary $\{X_t\}$ with this function as its acv.

Proof omitted here; we will revisit this result when we discuss Gaussian processes in more detail

[For any n, define $\Sigma_n = [\gamma(i-j)]_{i,j=1}^{n}$
- use this as variance covariance matrix]

Note that $\rho(\,)$ has the same properties as $\gamma(\,)$, except that

$$\rho(0) = 1.$$

EXAMPLE Let $\{X_t\}$ be defined by

$$X_t = A\cos(\omega t) + B\sin(\omega t)$$

where A and B are zero-mean, uncorrelated random variables with variance 1.

Then $E[X_t] = 0$

$$\begin{aligned}
\text{Cov}[X_t, X_{t+h}] &= E[X_t X_{t+h}] \\
&= E[A^2]\cos(\omega t)\cos(\omega(t+h)) \\
&\quad + E[B^2]\sin(\omega t)\sin(\omega(t+h)) \\
&= \cos(\omega h) \\
&= \cos(\omega(-h))
\end{aligned}$$

It follows that $g(h) = \cos(\omega h)$ is non-negative def.

EXAMPLE

Let

$$g(h) = \begin{cases} 1 & h = 0 \\ \rho & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

Comparison with an MA(1) process

$$\gamma(h) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \sigma^2\theta & h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

yields

$$\sigma^2(1+\theta^2) = 1$$

$$\sigma^2\theta = \rho$$

$$\therefore \quad \theta = \frac{1 \pm \sqrt{1-4\rho^2}}{2\rho}$$

with a <u>real</u> solution if $1 \geq 4\rho^2$

or $\frac{1}{2} \geq |\rho|$

If $|\rho| > \frac{1}{2}$, there is no MA(1) process with ACVF $g(h)$.

Now suppose $\rho > 1/2$. Let

$$G = [g(i-j)]_{ij}$$

be the $n \times n$ matrix formed by composing the $g(i-j)$ terms. Now let

$$\underset{\sim}{a} = (1, -1, 1, -1, \ldots, \ldots)^T$$

Then

$$\underset{\sim}{a}^T G \underset{\sim}{a} = n - 2(n-1)\rho < 0$$

if $n > \dfrac{2\rho}{2\rho - 1}$ $\therefore$ G is <u>not</u> non-neg def.

$\therefore$ $g(h)$ is not the ACVF of <u>any</u> stationary process if $\rho > 1/2$.

For $\rho < -1/2$, let

$$\underset{\sim}{a} = (1, 1, 1, \ldots)^T$$

Then again

$$\underset{\sim}{a}^T G \underset{\sim}{a} < 0$$

so G is not NND, and $g(h)$ is not the ACVF of a stationary process

$\therefore$ if

$$|\rho| > 1/2$$

g is not the ACVF of a stationary process.

Strict Stationarity

$\{X_t\}$ is strictly stationary if

$$(X_1, \ldots, X_n)^T \overset{\mathcal{L}}{=} (X_{1+h}, \ldots, X_{n+h})^T$$

for all $n, h \geq 0$.

$\overset{\mathcal{L}}{=}$ means "equal in law"

"has the same distribution as"

- if X_t are i.i.d then $\{X_t\}$ is strictly stationary
- if $\{X_t\}$ is strictly stationary, X_t are identically distributed
- if $\{X_t\}$ is strictly stationary, $\{X_t\}$ is weakly stationary if $E[X_t^2] < \infty$

Construction of Stationary Processes

Let $\{Z_t\}$ be an i.i.d. sequence, and define

$$X_t = g(Z_t, Z_{t-1}, \ldots, Z_{t-q})$$

for some real-valued function g.

X_t is a "filtered" version of $\{Z_t\}$

Now

$\{Z_t\}$ strictly stationary

$\Rightarrow \{X_t\}$ strictly stationary

and also X_s and X_t are <u>independent</u> if $|t-s| > q$

$\{X_t\}$ is "q-dependent"

Note also that

$$\gamma_X(h) = 0 \qquad |h| > q$$

A simple linear function g yields

MOVING AVERAGE

models

DEFINITION **The MA(q) Process**

$\{X_t\}$ is a moving-average process of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

where

$$\{Z_t\} \sim WN(0, \sigma^2)$$

$\theta_1, \dots, \theta_q$ are real constants

NOTE It can be shown that all zero mean, stationary, q-correlated processes can be represented as an MA(q) process.

NOTE The extension to $q = \infty$ yields an important representation for stationary processes

WOLD'S DECOMPOSITION

2.2 Linear Processes

$\{X_t\}$ is a **Linear Process** if, $\forall t$,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$

$\{\psi_j\}$ is a sequence of (real) constants s.t.

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

i.e $X_t = \psi(B) Z_t$

where

$$\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$$

A linear process is termed a

(non-anticipating) moving average

if $\psi_j = 0 \quad \forall j < 0$

i.e

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

we write $\{X_t\} \sim MA(\infty)$

NOTE $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures that

$\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ converges

in mean square

i.e.

$$X_t \overset{m.s}{=} \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

i.e $\forall t$,

$$\lim_{n \to \infty} E\left[\left(X_t - \sum_{j=-n}^{n} \psi_j Z_{t-j}\right)^2\right] = 0$$

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \Rightarrow \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$$

$$\Rightarrow \sum_{j=-n}^{n} |\psi_j| < \infty$$

NOTE

$$E[|Z_t|] \leq \sigma^2$$

$$\therefore E[|X_t|] \leq \sum_{j=-\infty}^{\infty} |\psi_j| E[|Z_{t-j}|]$$

$$\leq \sigma \sum_{j=-\infty}^{\infty} |\psi_j|$$

$$< \infty$$

We can conclude from this that

$$\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \text{ converges almost surely.}$$

$\psi(B)$ is a <u>linear filter</u> acting on the white noise series $\{Z_t\}$ returning stationary series $\{X_t\}$.

It transpires that $\psi(B)$ acting on <u>any</u> stationary series returns a stationary series.

<u>THEOREM</u> Let $\{Y_t\}$ be a stationary time series with mean zero and ACVF $\gamma_Y(.)$

Let $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$

with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Let $\{X_t\}$ be defined by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t$$

Then

$$E[X_t] = 0$$

and $\{X_t\}$ is stationary with acvf

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j)$$

If $\{X_t\}$ is a linear process, then

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2.$$

Proof $E[X_t] < \infty$ as

$$E[|X_t|] \leq \sum_{j=-\infty}^{\infty} |\psi_j| \sigma < \infty$$

and $E[Y_t] = 0$ yields $E[X_t] = 0$.

$$E[X_{t+h} X_t] = E\left[\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t+h-j}\right)\left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}\right)\right]$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k E\left[Y_{t+h-j} Y_{t-k}\right]$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h-j+k)$$

as all sums convergent.

If $\{Y_t\}$ is $WN(0, \sigma^2)$ then

$$\gamma_Y(h-j+k) = \begin{cases} \sigma^2 & h = j-k \\ 0 & \text{otherwise} \end{cases}$$

$$h = j-k \iff k = j-h$$

$$\therefore \quad E[X_{t+h} X_t] = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} \sigma^2$$

$$= \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2 .$$

NOTE If $\alpha(B)$ and $\beta(B)$ are two filters

$$\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$$

$$\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$$

such that $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty, \sum_{j=-\infty}^{\infty} |\beta_j| < \infty$

Then

$$\psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B)$$

is such that

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

and $\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}.$

Thus

$$X_t = \psi(B) Y_t$$

$$= \alpha(B)\beta(B) Y_t$$

is a stationary series if $\{Y_t\}$ is a stationary series

i.e. successive filters can be applied in turn to a stationary process.

Note that if $W_t = \beta(B) Y_t$ then $\{W_t\}$ is stationary, and

$$X_t = \alpha(B) W_t.$$

EXAMPLE AR(1) process

Let $\{Z_t\} \sim WN(0, \sigma^2)$, $|\phi| < 1$

and $\{X_t\}$ defined by

$$X_t = \phi X_{t-1} + Z_t \quad (E[X_t Z_s] = 0 \quad s > t)$$

Rewriting, $\{X_t\}$ is the solution of the equation

$$X_t - \phi X_{t-1} = Z_t$$

or $$(1 - \phi B) X_t = Z_t .$$

Consider the linear process

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

Note that $|\phi| < 1$

$$\Rightarrow \sum_{j=0}^{\infty} |\phi^j| < \infty$$

so the process satisfies the conditions of the earlier definition.

Now, by the recursion

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t \\ &= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t \\ &= \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\ &= \cdots = \lim_{n\to\infty} \left\{ \phi^n X_{t-n} + \sum_{j=0}^{n-1} \phi^j Z_{t-j} \right. \\ &= \sum_{j=0}^{\infty} \phi^j Z_{t-j} \end{aligned}$$

as $\phi^n \longrightarrow 0$ as $n \to \infty$

Also

$$\gamma_X(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma^2$$

$$= \frac{\sigma^2 \phi^h}{1-\phi^2} \qquad h \geq 0.$$

Now suppose that $\{Y_t\}$ is another process that solves the AR(1) equation and is stationary. Then by the above recursive argument

$$E\left[\left(Y_t - \sum_{j=0}^{k} \phi^j Z_{t-j}\right)^2\right] = \phi^{2k+2} E\left[Y_{t-k-1}^2\right]$$

$$\longrightarrow 0$$

as $k \to \infty$, as $E[Y_t^2] < \infty$ due to stationarity.

$$\therefore \quad Y_t \overset{m.s.}{=} \sum_{j=0}^{\infty} \phi^j Z_{t-j} \overset{m.s.}{=} X_t$$

and Y_t and X_t are equal in mean square.

Therefore X_t is the <u>unique</u> stationary solution to

$$X_t - \phi X_{t-1} = Z_t$$

for $|\phi| < 1$ with $\{Z_t\} \sim WN(0, \sigma^2)$

$\therefore$ for any stationary AR(1) process there is a unique MA(∞) representation, with coefficients

$$\psi_j = \phi^j \qquad j \geq 0$$

(and $j < 0$ gives $\psi_j = 0$).

NOTE If $|\phi| = 1$

i.e $X_t = X_{t-1} + Z_t$ or $X_t = -X_{t-1} + Z_t$.

Then $\{X_t\}$ is not stationary.

Suppose $\{X_t\}$ is a stationary solution. Then

$$X_t = \phi X_{t-1} + Z_t$$

$$= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t$$

$$= \vdots$$

$$= \phi^{n+1} X_{t-n-1} + \sum_{j=0}^{n} \phi^j Z_{t-j}$$

$$\text{[redacted]} = X_t - \phi^{n+1} X_{t-n-1}$$

Then

$$\mathrm{Var}[R_{t,n}] = \sum_{j=0}^{n} \phi^{2j} \mathrm{Var}[Z_{t-j}]$$

$$= (n+1)\sigma^2$$

$$\longrightarrow \infty$$

as $n \to \infty$.

$\therefore \mathrm{Var}[X_t] \longrightarrow \infty$

as $n \to \infty$: a contradiction.

$\therefore$ No stationary solution exists.

If $|\phi| > 1$

$$\sum_{j=0}^{\infty} \phi^j z_{t-j}$$

diverges.

But

$$X_t = \phi X_{t-1} + z_t$$

$$\Rightarrow X_t = \frac{1}{\phi} X_{t+1} - \frac{1}{\phi} z_{t+1}$$

$$= \frac{1}{\phi^2} X_{t+2} - \frac{1}{\phi} z_{t+1} - \frac{1}{\phi^2} z_{t+2}$$

$$= \ldots$$

$$\therefore \quad X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}$$

i.e $\{X_t\}$ depends on "future" values of $\{Z_t\}$

We will restrict attention to future independent ("causal", "non-anticipating") processes

<u>NOTE</u> $\quad X_t - \phi X_{t-1} = Z_t$

$$\therefore \quad (1-\phi B) X_t = Z_t$$

$$\Rightarrow \quad X_t = (1-\phi B)^{-1} Z_t = \pi(B) Z_t$$

$$= \left\{ \sum_{j=0}^{\infty} \phi^j B^j \right\} Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

2.3 General Autoregressive Processes

Let $Z_t \sim WN(0, \sigma^2)$

Definition AR(p) Process

Consider $\{X_t\}$ defined (as the solution to)

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t$$

This process is the Autoregressive Process of order p

Here $\phi_1, \dots, \phi_p$ are real constants

NOTE: We do not know from this definition whether $\{X_t\}$ is uniquely defined or stationary.

EXAMPLE AR(2)

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$$

Now $E[Z_t] = 0$, $Var[Z_t] = \sigma^2 < \infty$

$\therefore E[X_t] = 0$

To compute covariance, we consider

$$E[X_t X_{t+h}]$$

- not straightforward.

First, note that

$$(1 - \phi_1 B - \phi_2 B^2) X_t = Z_t .$$

$$\Rightarrow X_t = (1 - \phi_1 B - \phi_2 B^2)^{-1} Z_t$$

$$= \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

where ψ_j is the coefficient of B^j in the power series expansion of

$$(1 - \phi_1 B - \phi_2 B^2)^{-1}$$

Consider

$$\phi(z) = (1 - \phi_1 z - \phi_2 z^2)$$

$$= (1 - \xi_1 z)(1 - \xi_2 z) \quad \text{say}$$

say, where (ξ_1, ξ_2) are the complex conjugate pair

In general $(1 - \xi z)^{-1}$ has radius of convergence given by

$$|\xi| < 1 \quad \text{for} \quad |z| = 1$$

ξ_1, ξ_2: the "reciprocal roots" of

$$\phi(z) = 0$$

and we require $|\xi_1|, |\xi_2| < 1$

Now

$$\phi(B)X_t = Z_t \Rightarrow X_t = [\phi(B)]^{-1} Z_t$$

$$\Rightarrow X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

where ψ_j is the coefficient of B^j in the expansion

$$(1-\zeta_1 B)^{-1}(1-\zeta_2 B)^{-1}$$

Now

$$(1-\zeta_1 B)^{-1}(1-\zeta_2 B)^{-1} = \left(\sum_{j=0}^{\infty} \zeta_1^j B^j\right)\left(\sum_{k=0}^{\infty} \zeta_2^k B^k\right)$$

$$= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \zeta_1^j \zeta_2^k B^{j+k}$$

$\therefore$ Coefficient of B^r is

$$\sum_{k=0}^{r} \xi_1^k \xi_2^{r-k} = \psi_r$$

Now $|\psi_r| \leq \sum_{k=0}^{r} |\xi_1|^k |\xi_2|^{r-k}$

$$= \sum_{k=0}^{r} |\xi_2|^r$$

$$= (r+1) M^r$$

where $M = |\xi_2| < 1$

as $|\xi_1|^k = |\xi_2|^k$

$$\therefore \sum_{j=0}^{\infty} |\psi_j| \leq \sum_{j=0}^{\infty} (j+1) M^j$$

$$= \frac{d}{dM} \left\{ \sum_{j=0}^{\infty} M^{j+1} \right\}$$

$$= \frac{d}{dM} \left\{ \frac{M}{1-M} \right\}$$

$$= \frac{1}{(1-M)^2} < \infty$$

$$\therefore \quad X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

is a linear process, and is hence stationary.

This argument can be extended to the AR(p) case, writing

$$\phi(B) = \prod_{r=1}^{p} (1 - \xi_r B)$$

where $\xi_1, \ldots, \xi_p$ are the reciprocal roots of

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p = 0$$

All we require is that

$$|\xi_r| < 1$$

for $r = 1, \ldots, p$. If $\eta_r = \frac{1}{\xi_r}$

then $\eta_1, \ldots, \eta_r$ solve

$$\phi(z) = (z - \eta_1)(z - \eta_2) \ldots (z - \eta_p) = 0$$

with $|\eta_r| > 1$.

Therefore a sufficient condition for $\{X_t\}$ defined by

$$\phi(B)X_t = Z_t$$

to be stationary is that the p roots of the equation

$$\phi(z) = 0$$

denoted $\eta_1, \ldots, \eta_p$ are such that

$$|\eta_j| > 1$$

Using a similar trick to that used for the AR(1)

if $|m_j| < 1$ we can construct a non-causal representation that is also stationary.

Thus the AR(p) process is stationary if

$$|m_j| \neq 1$$

$\forall j$. In fact, using the uniqueness argument used for AR(1), we can deduce that $\{X_t\}$ is stationary only if $|m_j| \neq 1$.

Ex 1

$$X_t - \frac{13}{4} X_{t-1} + \frac{3}{4} X_{t-2} = Z_t$$

$$\therefore \phi(z) = 1 - \frac{13}{4} z + \frac{3}{4} z^2$$

$$= (1 - 3z)\left(1 - \frac{1}{4} z\right)$$

$\therefore$ Reciprocal roots (real!) are $3, 1/4$

(roots are $\frac{1}{3}, 4$)

$\therefore$ Non-stationary

$$\left(1 - \frac{1}{4} B\right)(1 - 3B) X_t = Z_t$$

But

$$(1 - 3B) X_t$$

has a diverging expansion.

NOTE If

$$\phi(z) = (1 - \xi_1 z)(1 - \xi_2 z)$$

has two real roots

$$|\psi_r| \leq \sum_{k=0}^{r} |\xi_1|^k |\xi_2|^{r-k}$$

$$\leq (r+1) M^r$$

where

$$M = \max\{|\xi_1|, |\xi_2|\}$$

and the same result follows.

Summary

MA processes: Always stationary

ACVs vanishes beyond lag q

i.e $\gamma(h) = 0 \quad |h| > q$

AR processes Stationary if ad only if roots of $\phi(z) = 0, \eta_1, \ldots, \eta_p$ have modulus $\neq 1$.

Causal if $|\eta_j| > 1$.

Linear / MA(∞) process representation.

2.4 ARMA PROCESSES

MA: $X_t = \theta(B) Z_t$

$= Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$

AR: $\phi(B) X_t = Z_t$

$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t$

DEFINITION ARMA PROCESS

$\{X_t\}$ is an ARMA(p,q) process if it is stationary and satisfies

$$\phi(B) X_t = \theta(B) Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$

EXAMPLE ARMA(1,1)

$\{X_t\}$ satisfies

$$\phi(B)X_t = \theta(B)Z_t \qquad (*)$$

$$\phi(B) = 1 - \phi B$$
$$\theta(B) = 1 + \theta B$$

Let $|\phi| < 1$, and define

$$\alpha(z) = \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \phi^j z^j$$

Then

$$X_t = \alpha(B)\theta(B)Z_t = \psi(B)Z_t$$

where $\psi(B) = (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B)$

$$\therefore \quad \psi_0 = 1$$

$$\psi_j = (\phi + \theta)\phi^{j-1} \qquad j \geq 1$$

As before, it follows that

$$X_t = Z_t + (\phi+\theta)\sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}$$

is the MA(∞) / linear process representation of the solution to the ARMA(1,1) eq^n (*)

For $|\phi| > 1$: using the same trick as for AR(1), the solution to (*) is

$$X_t = -\frac{\theta}{\phi} Z_t - (\theta+\phi)\sum_{j=1}^{\infty} \phi^{-(j+1)} Z_{t+j}$$

For $|\phi| < 1$: stationary, causal

$|\phi| > 1$: stationary, non-causal

$|\phi| = 1$: non-stationary

INVERTIBILITY For the ARMA(p,q) process

$$\phi(B) X_t = \theta(B) Z_t$$

Let $\beta(z) = \dfrac{1}{\theta(z)}$

If the power series expansion

$$\beta(z) = \sum_{j=0}^{\infty} \beta_j z^j$$

exists then we can rewrite

$$\beta(B)\,\phi(B) X_t = Z_t$$

When does the power series converge?

— by similar arguments to those for AR(p)

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q$$

$$= \prod_{i=1}^{q} (1 - \omega_i z)$$

$\omega_1, \ldots, \omega_q$ - "reciprocal" roots of the MA equation

$\therefore$ we need

$$|\omega_i| < 1 \qquad i = 1, \ldots q$$

or

$$|\lambda_i| > 1 \qquad i = 1, \ldots q$$

if $\lambda_i = \omega_i^{-1}$.

For ARMA(1,1),

$$\theta(B) = 1 + \theta B$$

$$\beta(B) = 1 - \theta B + \theta^2 B^2 - \ldots .$$

provided $|\theta| < 1$.

Then if $\pi(B) = \beta(B)\phi(B)$

we have

$$\pi(B) X_t = z_t$$

$$\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$$

where

$$\pi_0 = 1$$

$$\pi_j = -(\phi + \theta)(-\theta)^{j-1}$$

If $|\theta| > 1$, using the "time-reversal" trick

$$Z_t = -\frac{\phi}{\theta} X_t + (\phi+\theta)\sum_{j=1}^{\infty} (-\theta)^{j-1} X_{t+j}$$

That is, if $|\theta| \neq 1$ the ARMA(1,1) process has an AR(∞) representation

$|\theta| < 1$: stationary if $|\phi| \neq 1$
causal if $|\phi| < 1$
"invertible" if $|\theta| < 1$.

$|\theta| > 1$: stationary if $|\phi| \neq 1$
causal if $|\phi| < 1$
"non-invertible"

"invertible": Z_t can be expressed

$$Z_t = \pi(B)X_t = \sum_{j=-\infty}^{\infty} \pi_j X_{t-j}$$

where $\pi_j = 0$ if $j < 0$

and $$\sum_{j=0}^{\infty} |\pi_j| < \infty$$

"non-invertible": Z_t can be expressed in the same way, but with

$$\pi_j = 0 \text{ if } j > 0.$$

NOTE It is possible to convert a

non-causal, non-invertible

representation of $\{X_t\}$ w.r.t. $\{Z_t\} \sim WN(0, \sigma^2)$ to a

causal, invertible one

$$\phi^*(B) X_t = \theta^*(B) W_t$$

where

$\{W_t\} \sim WN(0, \sigma^2)$ also.

$\therefore$ we need only consider causal invertible ARMA(p,q) processes.

2.5 Moment Estimation for Stationary Processes

For an arbitrary stationary process $\{X_t\}$

To estimate $\mu = E[X_t]$ from a finite observation of the process $X_1, \ldots, X_n$, the moment estimator

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

can be used; $\overline{X}_n$ is unbiased and consistent

$$\overline{X}_n \xrightarrow{r=2} \mu$$

as $n \to \infty$, provided the variance of the estimator stays finite, and goes to zero as $n \to \infty$.

$$\mathrm{Var}[\bar{X}_n] = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}[X_i, X_j]$$

$$= \frac{1}{n^2} \sum_{h=-n}^{n} (n - |h|)\, \gamma_X(h)$$

$$= \frac{1}{n} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h)$$

Provided

$$\gamma(h) \to 0$$

as $h \to \infty$

$$\mathrm{Var}[\bar{X}_n] \longrightarrow 0$$

$$\bar{X}_n \xrightarrow{r=2} \mu$$

$$\bar{X}_n \xrightarrow{\text{m.s.}} \mu.$$

If $\{X_t\}$ is stationary, then

(i) $\bar{X}_n \xrightarrow{n\to\infty} \mu$ if $\gamma_x(n) \to 0$

(ii) $n\,\mathrm{Var}[\bar{X}_n] \xrightarrow{P} \sum_{|h|<\infty} \gamma_x(h)$

if $\sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty$.

NOTE It is not necessarily the case that

$$\sum_{n=-\infty}^{\infty} |\gamma_x(h)| < \infty$$

"LONG MEMORY" MODELS.

For Gaussian, linear, ARMA processes

$$\bar{X}_n \sim AN\left(\mu, \frac{1}{n} V_{\gamma_x}\right)$$

where

$$V_{\gamma_x} = \sum_{|h|<\infty} \gamma_x(h)$$

As usual, we estimate the asymptotic variance by some sample-based quantity

e.g $$\hat{V} = \sum_{h=-l}^{l} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}_n(h)$$

for some l, and estimate $\hat{\gamma}_n$ of γ_x.

Typically

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)$$

As before, we form the matrix of sample covariances

$$\hat{\Gamma}_{n,k} = \left[\hat{\gamma}_n(i-j)\right]_{i,j=1}^{k} \qquad k \geq 1.$$

which is <u>non-negative definite</u> (and is in fact positive definite / nonsingular often)

Also

$$\hat{R}_{n,k} = \frac{\hat{\Gamma}_{n,k}}{\hat{\gamma}_n(0)} \qquad k \geq 1$$

is the sample autocorrelation.

If we know the form of γ_X and ρ_X we can deduce the asymptotic variance / covariance matrix of $\hat{\gamma}_n$, $\hat{\rho}_n$

- this will allow testing of parametric (ARMA) models

For example, if $\hat{\rho}_n$ is the acfs estimator then

$$\hat{\rho}_n \sim AN\left(\rho_X, \frac{1}{n} W\right)$$

where $W = [w_{ij}]$.

where the w_{ij} are given by

BARTLETT'S FORMULA.

$$\omega_{ij} = \sum_{k=1}^{\infty} \left[\rho_X(k+i) + \rho_X(k-i) - 2\rho_X(i)\rho_X(k)\right] \left[\rho_X(k+j) + \rho_X(k-j) - 2\rho_X(j)\rho_X(k)\right]$$

EXAMPLES

1. $\{X_t\} \sim IID(0, \sigma^2)$: $\omega_{ij} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

2. $\{X_t\} \sim MA(1)$

$$X_t = Z_t + \theta Z_{t-1}$$

with $\{Z_t\} \sim WN(0, \sigma^2)$

Recall $\rho_X(h) = 0 \quad |h| > 1$.

After some algebra, the diagonal elements of W, w_{ii}, are

$$w_{ii} = \begin{cases} 1 - \rho_x^2(1) + 4\rho_x^4(1) & i = 1 \\ 1 + 2\rho_x^2(1) & i > 1 \end{cases}$$

Recall that for MA(1) with parameter θ.

$$\rho_x(1) = \frac{\theta}{1+\theta^2}$$

Thus $$\sqrt{n}\left(\hat{\rho}_x(i) - \rho_x(i)\right) \xrightarrow{d} N(0, w_{ii})$$

if $X_t \sim MA(1)$ with parameter θ.

3. AR(1) with $|\phi| < 1$.

If $\{X_t\} \sim AR(1)$ and $|\phi| < 1$

$$w_{ii} = \frac{(1-\phi^{2i})(1+\phi^2)}{1-\phi^2} - 2i\phi^{2i}$$

Thus if $\{X_t\} \sim AR(1)$ with parameter ϕ

$$\sqrt{n}\left(\hat{\rho}_n(i) - \rho_x(i)\right) \xrightarrow{d} N(0, w_{ii})$$

$\longrightarrow$ pontwise CIs at each lag.

Computation of W allows simultaneous CIs to be constructed.

2.6 Forecasting Stationary Processes

We aim to predict X_{n+h} given data: $X_1, X_2 \ldots, X_n$ observed to be $x_1, x_2, \ldots, x_n$.

We focus on *linear* predictors

- linear combnations of past values

without reference to model-based procedures

i.e we do not use distributional properties of $\{X_t\}$ or $\{Z_t\}$.

CRITERION : MINIMUM MSE

i.e we seek to choose constants

$$\underset{\sim}{a}_n = (a_0, a_1, \ldots, a_n)^T$$

to minimize the MSE

$$E\left[\left(X_{n+h} - \left\{a_0 + \sum_{i=1}^{n} a_i X_{n-i+1}\right\}\right)^2 \;\middle|\; \underset{\sim}{X} = \underset{\sim}{x}\right]$$

where $\underset{\sim}{X} = (X_1, \ldots, X_n)^T$

$\underset{\sim}{x} = (x_1, \ldots, x_n)^T$

We write

$$a_0 + \sum_{i=1}^{n} a_i X_{n-i+1} \equiv P_n X_{n+h}$$

i.e. P_n is the n-data prediction operator.

The problem of choosing $a_0, \ldots a_n$ to minimize MSE is a standard problem in statistics, solved using estimating equations (equivalent of the score equations used in model-based inference)

We need to solve (simultaneously)

$$E\left[X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n-i+1} \right] = 0$$

and

$$E\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n-i+1} \right) X_{n-j+1} \right] = 0$$

for $j = 1, \ldots, n$.

Re-writing this in vector notation

$$\Gamma_n \underset{\sim}{a}_n^{(0)} = \underset{\sim}{\gamma}_x(n,h)$$

where

$$\Gamma_n = \left[\gamma_x(i-j)\right]_{i,j=1}^{n}$$

$$\underset{\sim}{\gamma}_x(n,h) = (\gamma_x(h), \ldots, \gamma_x(h+n-1))^T$$

and

$$a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$$

$$\underset{\sim}{a}_n^{(0)} = (a_1, \ldots, a_n)^T$$

P_n is <u>uniquely</u> determined by this system

Denote the solution to

$$\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma}_x(n,h)$$

by

$$\underset{\sim}{a}_x(n,h) = (a_x(n,1), \ldots, a_x(n,h))^T$$

Then

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_x(n,i)(X_{n-i+1} - \mu)$$

<u>NOTE</u> By construction

$$E\left[(X_{n+h} - P_n X_{n+h})\right] = 0$$

and also, for $j = 1, \ldots, n$,

$$E\left[\left(X_{n+h} - P_n X_{n+h}\right) X_j\right] = 0$$

The minimum MSE is

$$E\left[\left(X_{n+h} - P_n X_{n+h}\right)^2\right]$$

$$= \gamma_X(0) - \left\{\underset{\sim}{a}_X(n,h)\right\}^T \underset{\sim}{\gamma}_X(n,h)$$

NOTE Solution of

$$\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma}_X(n,h)$$

is straightforward due to the form of Γ_n.

EXAMPLE $AR(1)$: $\{Z_t\} \sim WN(0, \sigma^2)$

$$X_t = \phi X_{t-1} + Z_t$$

$$\Gamma_n = \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & & \\ \phi^2 & \phi & 1 & & \\ \vdots & & & \ddots & \\ \phi^{n-1} & & & \phi & 1 \end{bmatrix}$$

$$\underset{\sim}{\gamma}_x(n,h) = (\phi, \phi^2, \phi^3, \ldots, \phi^n)^T$$

$$\therefore \underset{\sim}{a}_x(n,h) = (\phi, 0, \ldots, 0)^T$$

An intuitively reasonable result!

1.e

$$P_n X_{n+1} = \phi X_n$$
$$P_n X_{n+2} = \phi^2 X_n$$

etc.

NOTE Γ_n is an example of a TOEPLITZ MATRIX

$$\begin{bmatrix} a & b & c & & & & & & \\ b & a & b & c & & & & & \\ c & b & a & b & \ddots & & & & \\ & c & b & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \ddots & c & \\ & & & & \ddots & b & a & b & \\ & & & & & c & b & a & \end{bmatrix}$$

"constant diagonals"

General Prediction Considerations

Let $Y, W_1, \ldots, W_n$ be r.vs such that

$$E[Y] = \mu_Y$$
$$E[W_i] = \mu_i$$

$$\mathrm{Var}[Y], \mathrm{Var}[W_i] < \infty.$$

$$\mathrm{Cov}(Y, W_i), \mathrm{Cov}(W_i, W_j) < \infty.$$

Let $\underset{\sim}{W} = (W_n, \ldots, W_1)^T$

$$\underset{\sim}{\mu_W} = (\mu_n, \ldots, \mu_1)^T$$

Write

$$\underset{\sim}{\gamma} = \mathrm{Cov}(Y, \underset{\sim}{W}) \qquad n \times 1$$

$$= \left(\mathrm{Cov}(Y, W_n), \ldots, \mathrm{Cov}(Y, W_1)\right)^T$$

$$\Gamma = \mathrm{Cov}(\underset{\sim}{W}, \underset{\sim}{W}) \qquad n \times n$$

$$= \left[\mathrm{Cov}(W_{n-i+1}, W_{n-j+1})\right]_{i,j=1}^{n}$$

Then the best linear predictor of Y, given $W_1, \ldots, W_n$ can be found using identical methods

Write the best linear predictor

$$p(Y|\underset{\sim}{W}) = \mu_Y + \underset{\sim}{a}^T(\underset{\sim}{W} - \underset{\sim}{\mu_W})$$

where $\underset{\sim}{a}$ solves

$$\Gamma \underset{\sim}{a} = \underset{\sim}{\gamma}$$

MSE is

$$E\left[(Y - p(Y|\underset{\sim}{W}))^2\right] = Var[Y] - \underset{\sim}{a}^T\underset{\sim}{\gamma}$$

$P(Y|\underset{\sim}{W})$ converts $\underset{\sim}{W}$ into a prediction for Y

[a more general "prediction operator"]

PROPERTIES OF $P(Y|\underset{\sim}{W})$

Let $E[U^2], E[V^2] < \infty$

$\beta, \alpha_1, \ldots, \alpha_n$ be constants

Then

1. $P[U|\underset{\sim}{W}] = E[U] + \underset{\sim}{a}^T(\underset{\sim}{W} - \underset{\sim}{\mu}_W)$

where

$$\Gamma \underset{\sim}{a} = \underset{\sim}{\gamma}$$

2. $E[U - P(U|\underset{\sim}{W})] = 0$

$E[(U - P(U|\underset{\sim}{W})\underset{\sim}{W}] = 0$

3. $E[(U - P(U|\underset{\sim}{W}))^2] = Var[U] - \underset{\sim}{a}^T \underset{\sim}{\gamma}$

4. $P(\alpha_1 U + \alpha_2 V + \beta \,|\, \underset{\sim}{W})$

$= \alpha_1 P(U|\underset{\sim}{W}) + \alpha_2 P(V|\underset{\sim}{W}) + \beta$

(linear operator)

5. $P\left(\sum_{i=1}^{\hat{}} \alpha_i W_i + \beta \,|\, \underset{\sim}{W}\right) = \sum_{i=1}^{\hat{}} \alpha_i W_i + \beta$

(predicting the part)

6. $P(U|\underset{\sim}{W}) = E[U]$

if $\underset{\sim}{\gamma} \equiv 0$. (no prediction power if predictors are uncorrelated with U).

7. $P(U|\underset{\sim}{W})$

$$= P\left(P(U|\underset{\sim}{W}, \underset{\sim}{V})|\underset{\sim}{W}\right)$$

provided $E[\underset{\sim}{V}\underset{\sim}{V}^T]_{ij} < \infty$.

(iterated prediction)

These general results follow from general properties of r.v.s.

They allow us to <u>predict</u> for AR(p) processes

i.e. $\phi(B)X_t = Z_t \qquad Z_t \sim WN(0,\sigma^2)$

such that $\{X_t\}$ is stationary and causal.

Then it follows that

$$P_n X_{n+1} = \sum_{i=1}^{n} \phi_i X_{t-i} = \phi(B)X_t$$

as

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t$$

$$\therefore P_n X_t = P_n\left[\phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t\right]$$

$$= \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p}$$

using 4, 5, 6.

For optimal (in MSE) prediction we require the solution to

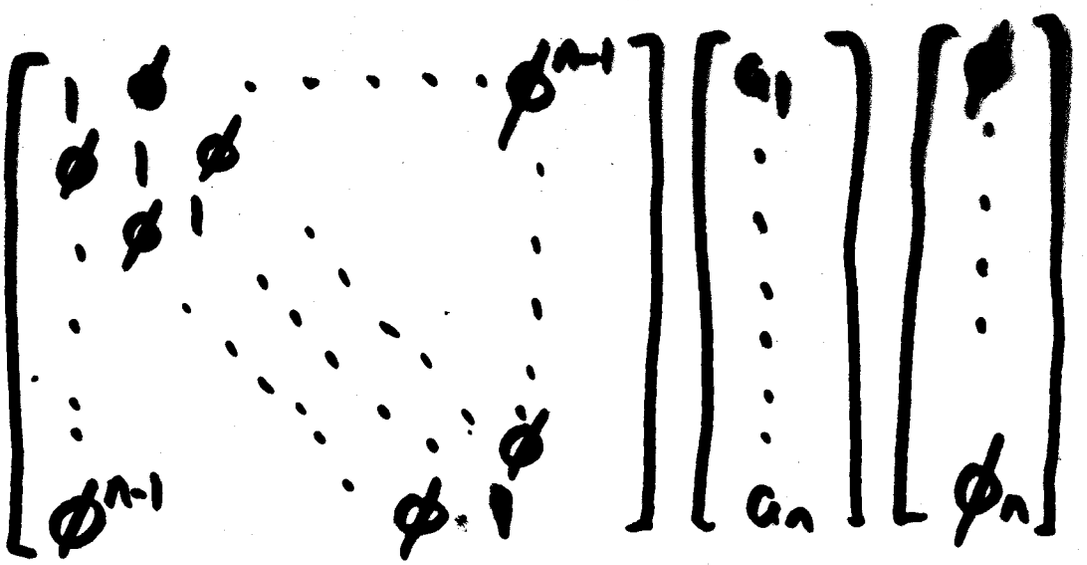
$$\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma}_x(n,h)$$

i.e.

$$\begin{bmatrix} \gamma_x(0) & \cdots & \gamma_x(n-1) \\ & \ddots & \\ \gamma_x(n-1) & & \gamma_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \gamma_x(h) \\ \vdots \\ \gamma_x(h+n-1) \end{bmatrix}$$

TOEPLITZ + SYMMETRIC

n equations in n variables

eg 

for AR(1).

If Γ_n is nonsingular,

$$\underset{\sim}{a} = \Gamma_n^{-1} \underset{\sim}{\gamma}_x(1,h)$$

If n is large, computing Γ_n^{-1} may be computationally demanding or inaccurate/unstable numerically.

Sufficient conditions for nonsingularity

(i) $\gamma_x(0) > 0$

(ii) $\gamma_x(h) \longrightarrow 0$ as $h \longrightarrow \infty$

These conditions are met for most stationary processes of interest

<u>NOTE</u>: This result relates to

$\gamma_x(h)$

not $\hat{\gamma}_x(h)$, but under these conditions

$\hat{\gamma}_x(h) \longrightarrow \gamma_x(h)$

so $\hat{\Gamma}_n$ is non singular w.p. 1 for n large enough.

Two recursive methods are used to compute the predictions sequentially

$P_n X_{n+1}$

↓

$P_{n+1} X_{n+2}$

↓

$P_{n+2} X_{n+3}$

etc

1. LEVINSON-DURBIN ALGORITHM
2. INNOVATIONS ALGORITHM.

LEVINSON-DURBIN ALGORITHM

Suppose we write

$$P_n X_{n+1} = \underset{\sim}{\phi}_n^T \underset{\sim}{X}_n = \sum_{j=1}^{n} \phi_{nj} X_{n-j+1}$$

where

$$\underset{\sim}{\phi}_n = \Gamma_n^{-1} \underset{\sim}{\gamma}_x(n, 1)$$

Let

$$v_n = E\left[(X_{n+1} - P_n X_{n+1})^2\right]$$

$$= \gamma_x(0) - \underset{\sim}{\phi}_n^T \underset{\sim}{\gamma}_x(n, 1)$$

be the prediction one-step-ahead variance.

RECURSIONS We compute $\underset{\sim}{\phi}_n$ from $\underset{\sim}{\phi}_{n-1}$:

Set

$$\phi_{n,n} = \frac{1}{v_{n-1}} \left[\gamma_x(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\, \gamma_x(n-j) \right]$$

$$\underset{\sim}{\phi}_n^{(n)} = \begin{pmatrix} \underset{\sim}{\phi}_{n-1} - \phi_{n,n}\, \underset{\sim}{\phi}_{n-1}^{R} \end{pmatrix}$$

where

$$\underset{\sim}{\phi}_n^{(n)} = (\phi_{n,1}, \ldots, \phi_{n,n-1})^T$$

$$\underset{\sim}{\phi}_{n-1}^{R} = (\phi_{n-1,n-1}, \ldots, \phi_{n-1,1})^T$$

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

Initialization: $\phi_{11} = \rho_x(1) = \dfrac{\gamma_x(1)}{\gamma_x(0)}$

$$v_0 = \gamma_x(0)$$

Proof By direct computation, and an inductive argument:

Let $R_n = \dfrac{1}{\gamma_x(0)} \Gamma_n$

$$\underset{\sim}{\rho}_x(n) = \frac{\underset{\sim}{\gamma}_x(n,1)}{\gamma_x(0)}$$

$$= (\rho_x(1), \ldots, \rho_x(n))^T$$

and show that $R_n \underset{\sim}{\phi}_n = \underset{\sim}{\rho}_x(n)$

$$\Rightarrow R_{n+1} \underset{\sim}{\phi}_{n+1} = \underset{\sim}{\rho}_x(n+1)$$

n=1 $\phi_{11} = \dfrac{\gamma_x(1)}{\gamma_x(0)} \quad \therefore \; R_1 \underset{\sim}{\phi}_1 = \underset{\sim}{\rho}_x(1)$

$$v_0 = \gamma_x(0)$$

Assume true for $n=k$.

n=k+1

$$R_{k+1}\,\underset{\sim}{\phi}_{k+1} = \begin{bmatrix} R_k & \underset{\sim}{\rho}_x^R(k) \\ \left[\underset{\sim}{\rho}_x^R(k)\right]^T & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\phi}_k - \phi_{k+1,k+1}\,\underset{\sim}{\phi}_k^R \\ \phi_{k+1,k+1} \end{bmatrix}$$

$$= \underset{\sim}{\rho}_x(k+1)$$

$$\left[R_k \underset{\sim}{\phi}_k = \underset{\sim}{\rho}_x(k) \quad \therefore \; R_k \underset{\sim}{\phi}_k^R = \underset{\sim}{\rho}_x^R(k) \right]$$

NOTE R_n is a symmetric, positive definite (non-singular) matrix

$\therefore$ R_n can be written

$$R_n = L_n L_n^T$$

where L_n is a lower triangular matrix

"LU decomposition"

"Cholesky decomposition"

These methods do not exploit the Toeplitz nature of R_n, and can be slower.

NOTE Define function $\alpha_x(n)$ $n=1,2,\ldots$ by

$$\alpha_x(0) = 1$$

$$\alpha_x(n) = \phi_{n,n} \qquad n=1,2,\ldots$$

where $\phi_{n,n}$ is found during the Levinson-Durbin recursion.

$\alpha_x(n)$ is termed the

"PARTIAL AUTOCORRELATION FUNCTION." PACF

NOTE $v_n = (1-\phi_{n,n}^2)\, v_{n-1}$

so v_{n-1} decreases to v_n by $(1-\phi_{n,n}^2)$ when X_n is included in the predictor

NOTE α_x is uniquely determined by γ_x

- thus α_x may be useful in model identification

- In ARMA models α_x is ecsily computed, and provides an alternative characterization to γ_x.

- Parameterization of a stationary model via α_x may be more straightforward.

Note The PACF is equivalently defined as the correlation between X_t and X_{t+n} after the "influence" of

$$X_{t+1}, \ldots, X_{t+n-1}$$

has been removed

i.e.

$$\alpha_x(n) = \phi_{n,n}$$

$$= \text{Corr}\left[(X_t - P(X_t | X_{t+1}, \ldots, X_{t+n-1})) \right.$$

$$\left. (X_{t+n} - P(X_{t+n} | X_{t+1}, \ldots X_{t+n}\right.$$

where $P(\cdot | \cdot)$ is the best linear predictor operator.

PARTIAL CORRELATION

For vector RVs $\underset{\sim}{X}, \underset{\sim}{Y}, \underset{\sim}{Z}$ the "partial correlation" of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ given $\underset{\sim}{Z}$ is defined by

$$\text{Corr}\left[(\underset{\sim}{X} - E[\underset{\sim}{X}|\underset{\sim}{Z}]), (\underset{\sim}{Y} - E[\underset{\sim}{Y}|\underset{\sim}{Z}]) \right]$$

i.e the correlation (matrix) formed when correlating the residuals from a regression of $\underset{\sim}{X}$ on $\underset{\sim}{Z}$ and the residuals from a regression of $\underset{\sim}{Y}$ on $\underset{\sim}{Z}$

For time series $\{X_t\}$.

$$\underset{\sim}{X} \equiv X_t$$

$$\underset{\sim}{Y} \equiv X_{t+n}$$

$$\underset{\sim}{Z} = (X_{t+1}, \ldots, X_{t+n-1})$$

and "regression" equates to "best linear prediction".

<u>NOTE</u> If $\{X_t\} \sim AR(p)$, then for $h > p$

$$P(X_{t+h} \mid X_{t+1}, \ldots, X_{t+h-1})$$

$$= \phi_1 X_{t+h-1} + \cdots + \phi_p X_{t+h-p}$$

$$\therefore X_{t+h} - P(X_{t+h} \mid X_{t+1}, \ldots, X_{t+h-1})$$

$$= Z_{t+h}$$

$$X_t - P(X_t \mid X_{t+1}, \ldots, X_{t+h-1}) = Z_t$$

$$\therefore \text{Corr}[Z_t, Z_{t+h}] = 0$$

$\therefore$ PACF is <u>zero</u> for $h > p$.

This is characteristic of the AR(p)

[Recall that for MA(q), ACF is zero if $h > q$]

The Innovations Algorithm

Suppose $\{X_t\}$ has mean zero

$$E[|X_t|^2] < \infty \quad \forall t.$$

Denote $\mathrm{Cov}[X_t, X_s] = E[X_t X_s]$

$$= K(t,s) \text{ say.}$$

Let $\hat{X}_n$ denote the one step prediction errors

$$\hat{X}_n = \begin{cases} 0 & n = 1 \\ P_{n-1} X_n & n = 2, 3, \ldots \end{cases}$$

$$V_n = E\left[(X_{n+1} - P_n X_{n+1})^2\right]$$

Let

$$U_n = X_n - \hat{X}_n$$

denote the prediction errors or "innovations".

ad $\underset{\sim}{U}_n = (U_1, \ldots, U_n)^T$

$\underset{\sim}{X}_n = (X_1, \ldots, X_n)^T$

Then

$$\underset{\sim}{U}_n = A_n \underset{\sim}{X}_n$$

where A_n is an $n \times n$ lower triangular matrix.

$$A_n = \begin{bmatrix} 1 & & & \\ a_{11} & 1 & & \\ a_{22} & a_{21} & 1 & \\ \vdots & \vdots & \ddots & \ddots \\ a_{n-1,n-1} & \cdots & \cdots & 1 \end{bmatrix} = [a_{ij}]$$

For a stationary process

$$a_{ij} = -a_j \qquad j = 1, \ldots, n$$

found by solving

$$\Gamma_n \underset{\sim}{a} = \underset{\sim}{\gamma}_x(n, 1)$$

as before.

$$\therefore \quad \hat{\underset{\sim}{X}}_n = \underset{\sim}{X}_n - \underset{\sim}{U}_n$$

$$= A_n^{-1} \underset{\sim}{U}_n - \underset{\sim}{U}_n = C_n \underset{\sim}{U}_n - \underset{\sim}{U}_n$$

$$= \Theta_n (\underset{\sim}{X}_n - \hat{\underset{\sim}{X}}_n) \quad \text{say}$$

where

$$C_n = A_n^{-1} = \begin{bmatrix} 1 & & & \\ \theta_{11} & 1 & & 0 \\ \theta_{22} & \theta_{11} & \ddots & \\ \vdots & & \ddots & \\ \theta_{n-1,n-1} & \cdots & \cdots & 1 \end{bmatrix}$$

$$\Theta_n = C_n - I_n$$

I_n — $n \times n$ identity.

as $\underset{\sim}{\hat{X}}_n = C_n (\underset{\sim}{X}_n - \underset{\sim}{\hat{X}}_n)$

Thus for the one step ahead prediction

$$\hat{X}_{n+1} = \begin{cases} 0 \\ \sum_{j=1}^{\hat{n}} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) \end{cases}$$

i.e the forecast for X_{n+1} can be constructed from the previous prediction errors once the matrix Θ_n has been computed.

Algorithm

Initialize : $V_0 = K(1,1) = E[X_1^2]$

Recursion : For each n, and $0 \le k < n$

$$\theta_{n,n-k} = \frac{1}{V_k}\left(K(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\,\theta_{n,n-j}\,V_j\right)$$

and $V_n = K(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2\,V_j$

THUS

LEVINSON-DURBIN

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_{n+1-j}$$

INNOVATIONS

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{n,j} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right)$$

LEVINSON-DURBIN - "NON CENTERED"

INNOVATIONS - "CENTERED"

If $D_n = X_n - \hat{X}_n$, then

$\{D_n\}$ is uncorrelated.

AR/MA prediction :

LEVINSON-DURBIN : better for AR(p)

as $\phi_{n,j} = 0$ for $n-j > p$

INNOVATIONS : better for MA(q)

as $\theta_{n,j} = 0$ for $n-j > q$

Both can be programmed as $O(n^2)$ procedures

For h-step ahead prediction:

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right)$$

where the θs are found by running the innovations algorithm forward h steps.

- follows as

$$P_n X_{n+h} = P_n \hat{X}_{n+h}$$

$$V_{n+h} = K(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta^2_{n+h-1,j} \, V_{n+h-j-1}$$

Forecasting using past values

Given $X_m, X_{m+1}, \ldots, X_{-2}, X_{-1}, X_0, X_1, \ldots X_n$ $(m<0)$ we construct the best linear predictor of X_{n+h} as

$$P_{m,n} X_{n+h} = \sum_{j=m}^{n} \pi_j X_j .$$

and

$$\tilde{P}_n X_{n+h} = \lim_{m \to -\infty} P_{m,n} X_{n+h}$$

We seek the coefficients $\{\pi_j\}$.

Using similar logic to the previous construction we seek to solve

$$E\left[\left(X_{n+h} - \tilde{P}_n X_{n+h}\right) X_{n+1-i}\right] = 0$$

simultaneously for $i = 1, 2, \ldots$

Suppose that a valid solution takes the form above. Then

$$E\left[\left(X_{n+h} - \sum_{j=1}^{\infty} \pi_j X_{n+1-j}\right) X_{n+1-i}\right] = 0$$

for $i = 1, 2, \ldots$

Thus

$$\sum_{j=1}^{\infty} \gamma_x(i-j)\pi_j = \gamma_x(h+i-1) \quad i = 1, 2, \ldots$$

Provided the RHS converges this system (uniquely) determines $\pi_1, \pi_2, \ldots$

This is a useful representation for

- AR
- MA
- ARMA

processes.

Thus if γ_x is completely specified we can compute the π_j

...

Now, from the infinite MA rep.

$$\tilde{P}_n X_{n+1} = \tilde{P}_n Z_{n+1} + \tilde{P}_n \left((\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{n+1-j} \right)$$

But $\tilde{P}_n Z_{n+1} = E[Z_{n+1}] = 0$

as Z_{n+1} is <u>uncorrelated</u> with $\{X_t\}$.

Also, as Z_s is <u>known</u> for $s \leq n$.

$$\tilde{P}_n Z_s = Z_s \quad \text{for} \quad s \leq n.$$

and $\tilde{P}_n$ is a linear operator.

EXAMPLE ARMA(1,1)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$$

with $\{Z_t\} \sim WN(0, \sigma^2)$

Suppose $\{X_t\}$ is causal and invertible

$|\phi|, |\theta| < 1$.

Then

MA : $X_{n+1} = Z_{n+1} + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{n+1-j}$

AR : $Z_{n+1} = X_{n+1} - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{n+1-j}$

$$\therefore \tilde{P}_n\left((\phi+\theta)\sum_{j=1}^{\infty}\phi^{j-1}Z_{n+1-j}\right)$$

$$= (\phi+\theta)\sum_{j=1}^{\infty}\phi^{j-1}\tilde{P}_n Z_{n+1-j}$$

$$= (\phi+\theta)\sum_{j=1}^{\infty}\phi^{j-1}Z_{n+1-j}$$

i.e $$\tilde{P}_n X_{n+1} = (\phi+\theta)\sum_{j=1}^{\infty}\phi^{j-1}Z_{n+1-j} \quad (1)$$

From the infinite AR rep.

$$\tilde{P}_n X_{n+1} = (\phi+\theta)\sum_{j=1}^{\infty}(-\theta)^{j-1}X_{n+1-j} \quad (2)$$

Combining ① + ② with the infinite representations

$$X_{n+1} - \tilde{P}_n X_{n+1} = Z_{n+1}$$

$$\therefore \quad E\left[(X_{n+1} - \tilde{P}_n X_{n+1})^2\right] = E\left[Z_{n+1}^2\right] = \sigma^2$$

∴ Under the best linear prediction using the infinite part, the MSE is σ^2.

Similar calculations work for pure MA/AR.

DETERMINISTIC / NON-DETERMINISTIC PROCESSES

Any process $\{X_t\}$ where, for each n

$$X_n - \tilde{P}_{n-1} X_n = 0$$

(in mean square, that is

$$E\left[(X_n - \hat{P}_{n-1} X_n)^2\right] = 0$$

for each n) is termed

(PURELY) DETERMINISTIC

The process is NON DETERMINISTIC if this does not hold.

The Wold Decomposition

If $\{X_t\}$ is stationary and non deterministic then $\{X_t\}$ can be represented via the sum of a linear combination of

- a white noise process $\{Z_t\}$
- a purely deterministic process $\{V_t\}$

i.e. for each t

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$$

where

(i) $\psi_0 = 1, \quad \sum_{j=1}^{\infty} \psi_j^2 < \infty$

(ii) $\{Z_t\} \sim WN(0, \sigma^2)$

(iii) $\{Z_t\}$ and $\{V_t\}$ are uncorrelated

(iv) $Z_t = \tilde{P}_t Z_t$ for all t.

(v) $V_t = \tilde{P}_s V_t$ for all s, t.

(vi) $\{V_t\}$ is deterministic

This representation is <u>unique</u>, and in fact

$$Z_t = X_t - \tilde{P}_{t-1} X_t \qquad \text{for each } t.$$

$$\psi_j = \frac{E[X_t Z_{t-j}]}{E[Z_t^2]} \qquad j = 1, 2, \ldots$$

$$V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j} \qquad \text{for each } t$$

If $\{V_t\} \equiv 0 \quad \forall t$, then X_t is purely non-deterministic

Thus the Wold decomposition decomposes the variation of $\{X_t\}$ into

- a deterministic part $\{V_t\}$ ("predictable")
- a non-deterministic part based on $\{Z_t\}$ and $\{\psi_j\}$ $j = 1, 2, \ldots$

3. ARMA MODELS

Recall that an ARMA(p,q) process is written

$$\phi(B) X_t = \theta(B) Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_p z^p$$

We recall the factorizations

$$\phi(z) = \prod_{j=1}^{p} (1 - \xi_j z)$$

$$\theta(z) = \prod_{j=1}^{q} (1 - \omega_j z)$$

We assume that $\phi(z)$ and $\theta(z)$ have no common factors

- if $\xi_j = \omega_k$ say some j, k.

  Then these terms would cancel in the representation as

$$(1 - \xi_j z) = (1 - \omega_k z)$$

$\therefore$ would become an ARMA$(p-1, q-1)$

We examine the specific form of

- the representations of these processes
- ACF / PACF
- Prediction operators.

3.1 Basic Properties

Recall that $\{X_t\}$ is stationary if and only if $|\xi_j| \neq$ $\forall j = 1, \ldots, p$

i.e if

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0$$

for all $|z| = 1$.

In fact:

$\{X_t\}$ is the unique stationary solution to

$$\phi(B) X_t = \theta(B) Z_t.$$

3.1.3 CAUSALITY AND INVERTIBILITY

$\{X_t\} \sim$ ARMA(p,q) is **CAUSAL** if there exist constants $\psi_0, \psi_1, \ldots$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

for all t.

i.e.

$$\phi(z) \neq 0 \quad \forall\, |z| \leq 1$$

i.e $|\xi_j| > 1 \quad \forall j = 1, \ldots, p.$

Note by using the power series expansions equating coefficients

$$\psi(z) = \frac{\theta(z)}{\phi(z)}$$

We obtain

$$\psi_j = \theta_j + \sum_{k=1}^{p} \phi_k \psi_{j-k} \qquad j = 0, 1, 2, \ldots$$

- for $j = 0, 1, \ldots$ we can compute the coefficients recursively.

$$(1 - \phi_1 z - \cdots - \phi_p z^p)(\psi_0 + \psi_1 z + \ldots + \psi_j z^j + \ldots)$$

$$= 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q$$

$\{X_t\} \sim$ ARMA(p,q) is **INVERTIBLE**

if there exist constants $\pi_0, \pi_1, \ldots$

such that

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \forall t$$

where $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

i.e. if $\theta(z) \neq 0 \quad \forall |z| \leq 1$

The recursions for $\{\pi_j\}$ are

$$\pi_j = -\phi_j - \sum_{k=1}^{q} \theta_k \pi_{j-k}.$$

$[\phi_0 \equiv -1]$

3.1.2 ACF and PACF for ARMA(p,q)

From previous results for linear processes if $\{X_t\} \sim ARMA(p,q)$

$$\gamma_X(h) = E[X_t X_{t+h}]$$

$$= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \qquad h = 0, \pm 1, \ldots$$

where σ^2 is the WN variance, and $\{\psi_j\}$ are the coefficients in the MA(∞) representation

$$X_t = \sum_{j=0}^{\infty} \psi_j z_{t-j}$$

EXAMPLES 1. ARMA(1,1) with $|\phi| < 1$.

$$\gamma_X(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

$$= \sigma^2 \left[\psi_0^2 + \sum_{j=1}^{\infty} \psi_j^2 \right]$$

$$= \sigma^2 \left[1 + (\theta+\phi)^2 \sum_{j=1}^{\infty} \phi^{2j} \right]$$

$$= \sigma^2 \left[1 + \frac{(\theta+\phi)^2}{1-\phi^2} \right]$$

$$\gamma_X(1) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1}$$

$$= \sigma^2 \left[(\theta+\phi) + (\theta+\phi)^2 \phi \sum_{j=1}^{\infty} \phi^{2j} \right]$$

$$= \sigma^2 \left[(\theta+\phi) + \frac{(\theta+\phi)^2 \phi}{1-\phi^2} \right]$$

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

$$= \phi^{h-1} \gamma_X(h) \qquad h \geq 2.$$

2. MA(q) [i.e $p=0$]

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & h = 0, 1, \ldots, q \\ 0 & \text{otherwise} \end{cases}$$

with $\gamma_X(-h) = \gamma_X(h)$

Alternate computation is direct from form of $\{X_t\}$

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p}$$

$$= Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

i.e.

$$\sum_{j=0}^{p} \phi_j X_{t-j} = \sum_{j=0}^{q} \theta_j Z_{t-j}$$

$\phi_0 = \theta_0 = 1.$

Premultiply by X_{t-k}

$$\sum_{j=0}^{p} \phi_j X_{t-k} X_{t-j} = \sum_{j=0}^{q} \theta_j X_{t-k} Z_{t-j}$$

Then take expectations on both sides

$$\sum_{j=0}^{p} \phi_j \gamma_x(k) = \sum_{j=0}^{q} \theta_j E[X_{t-k} Z_{t-j}]$$

$k=0$

$$E\left[X_t Z_{t-j}\right] = E\left[\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i}\right) Z_{t-j}\right]$$

$$= \psi_j E\left[Z_{t-j}^2\right]$$

$$= \psi_j \sigma^2$$

General k

$$E\left[X_{t-k} Z_{t-j}\right] = E\left[\left(\sum_{i=0}^{\infty} \psi_i Z_{t-k-i}\right) Z_{t-j}\right]$$

$$= \psi_{j+k} \sigma^2 .$$

$$\therefore \sum_{j=0}^{q} \theta_j E\left[X_{t-k} Z_{t-j}\right]$$

$$= \begin{cases} \sigma^2 \sum_{j=0}^{m} \theta_{k+j} \psi_j & 0 \le k < m \\ 0 & k \ge m \end{cases}$$

where

$$m = \max(p, q+1)$$

ψ_j = coefficient in MA(∞)

$\theta_0 = 1$.

∴ We wish to solve

$$\sum_{j=0}^{p} \phi_j \gamma_x(k-j) = \begin{cases} \sigma^2 \sum_{j=0}^{\infty} \theta_{m+j} \psi_j & 0 \le k < m \\ 0 & k \ge m \end{cases}$$

simultaneously, for γ_x, for each $k-j$ value.

For $k \ge m$

$$\gamma_x(k) - \phi_1 \gamma_x(k-1) - \dots - \phi_p \gamma_x(k-p) = 0.$$

$$\gamma_x(k+1) - \phi_1 \gamma_x(k) - \dots \phi_p \gamma_x(k-p+1) = 0$$

$$\vdots$$

Solution is known to be of the form

$$\gamma_X(h) = \sum_{j=1}^{p} \alpha_j \xi_j^h \quad \text{for } h \geq m-p$$

where $\xi_1^{-1}, \ldots, \xi_p^{-1}$ are the roots of the equation

$$\phi(z) = 0$$

as before.

For $0 \leq h \leq m-p$, we may use the equations (a linear system).

$$\sum_{j=0}^{p} \phi_j \gamma_X(k-j) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j} \psi_j$$

The Autocovariance Generating Function

The ACVGF is another tool for computing ACV sequences

Recall for real sequence $\{a_n\}$, the generating function related to this sequence is

$$G_a(z) = \sum_n a_n z^n$$

for $z \in \mathbb{C}$, with z in some annulus $\frac{1}{r} < |z| < r$

with $r \in \mathbb{R}$, and $r > 1$.

We compute the ACVGF, which is the generating function for the autocovariance sequence.

Suppose $\{X_t\}$ is defined by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j z_{t-j}$$

with $\{z_t\} \sim WN(0, \sigma^2)$

Suppose $\exists\ r > 1$ s.t. the sum

$$\sum_{j=-\infty}^{\infty} |\psi_j| z^j \qquad \frac{1}{r} < |z| < r$$

$$z \in \mathbb{C}$$

■ convergent.

Define the ACVGF, $G_X(z)$, by

$$G_X(z) = \sum_{h=-\infty}^{\infty} \gamma_X(h) z^h$$

for $z \in \mathcal{A}_r = \{z : r^{-1} < |z| < r\}$.

From previous work

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t]$$

$$= \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|}$$

$$\therefore \quad G_x(z) = \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|} z^h$$

$$= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{h=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} (z^h + z^{-h}) \right]$$

$$= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j z^j \right] \left[\sum_{h=-\infty}^{\infty} \psi_h z^{-h} \right]$$

$$= \sigma^2 \psi(z) \psi(z^{-1})$$

where

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$$

Hence if

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

Then

$$G_X(z) = \sigma^2 \psi(z) \psi(z^{-1})$$

where

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$$

This holds for any linear process for which the ACVGF exists (i.e when the annulus of convergence can be found).

ACVGF For ARMA(p, q)

$$\phi(B) X_t = \theta(B) Z_t .$$

For a stationary ARMA process, where

$$\phi(z) \neq 0$$

for $|z| = 1$, we have

$$\phi(B) X_t = \theta(B)$$

$$\Rightarrow X_t = \phi^{-1}(B)\theta(B) Z_t$$

$$= \psi(B) Z_t$$

where $\psi(z) = \dfrac{\theta(z)}{\phi(z)}$

Note that

$$\phi(z) \neq 0$$

when $|z| = 1$ by assumption

Hence

$$\psi(z) = \frac{\theta(z)}{\phi(z)}$$

is well defined and finite for z st

$$\frac{1}{r} < |z| < r$$

for some $r > 1$.

Thus

$$G_X(z) = \sigma^2 \psi(z)\psi(z^{-1})$$

$$= \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}$$

This gives a neat computation formula for the acv sequence of the ARMA process.

Also, recall that

$$\theta(z) = \prod_{j=1}^{q} (1 - w_j z)$$

$$\phi(z) = \prod_{j=1}^{p} (1 - \xi_j z)$$

NOTE If $\{Z_t\} \sim WN(0, \sigma^2)$

$$G_Z(z) = \sum_{j=-\infty}^{\infty} \gamma_Z(j) z^j$$

$$= \sigma^2$$

i.e constant, independent of z.

The ACVGF is constant if and only if the process is white noise.

[i.e only $\gamma_Z(0)$ is non-zero]

However, the ACVGF is not, in general, unique i.e we can find different processes with the same ACVGF.

Converting Non-causal / Non-invertible Processes

We can restrict attention to causal and invertible ARMA(p,q) processes as we are able to transform to a causal / invertible (C.I.) process even if the original process is non-causal / non-invertible (N.C.N.I).

We can also transform the

- parameters
- acv sequence.

EXAMPLE $\{X_t\} \sim MA(1)$

$$X_t = Z_t - 2Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$

Then the MA polynomial is

$$\theta(z) = 1 - 2z \qquad z \in \mathbb{C}$$

and $|z| > 1$, so the process is non-invertible.

Let

$$Z_t^* = (1 - \tfrac{1}{2}B)^{-1}(1 - 2B) Z_t$$

define a new process $\{Z_t^*\}$, with

$$(1 - \tfrac{1}{2}B) Z_t^* = (1 - 2B) Z_t = X_t.$$

1.e

$$\phi(B) z_t^* = \theta(B) z_t .$$

where

$$\phi(z) = 1 - \tfrac{1}{2} z$$

$$\theta(z) = 1 - 2z .$$

$\therefore$

$$z_t^* = \psi(B) z_t$$

where

$$\psi(z) = \frac{\theta(z)}{\phi(z)} = \frac{(1-2z)}{(1-\frac{1}{2}z)}$$

Then

$$Z_t^* = (1 - \tfrac{1}{2}B)^{-1} X_t$$

$$= \sum_{j=0}^{\infty} (\tfrac{1}{2})^j X_{t-j}$$

and the ACVGF of $\{Z_t^*\}$ is

$$G_{Z^*}(z) = \sigma^2 \psi(z) \psi(z^{-1})$$

where

$$\psi(z) = \frac{(1-2z)}{(1-\frac{1}{2}z)}$$

$$\psi(z^{-1}) = \frac{(1-2z^{-1})}{(1-\frac{1}{2}z^{-1})}$$

$$\therefore \Psi(z)\Psi(z^{-1}) = \frac{(1-2z)(1-2z^{-1})}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

$$= \frac{4(1-2z)(1-2z^{-1})}{(1-2z)(1-2z^{-1})}$$

$$= 4$$

as $$1-\frac{1}{2}z^{-1} = \frac{z^{-1}}{2}(2z-1)$$

$$1-\frac{1}{2}z = \frac{z}{2}(2z^{-1}-1)$$

$$\therefore G_{z^*}(z) = 4\sigma^2$$

$$\therefore \{z_t^*\} \sim WN(0, 4\sigma^2)$$

$$\therefore X_t = (1 - \tfrac{1}{2}B) z_t^* = \phi(B) z_t^*$$

$$= z_t^* - \tfrac{1}{2} z_{t-1}^*$$

has an <u>invertible</u> representation via

$\{z_t^*\}$

A similar result holds for

- noncausal → causal AR processes
- ARMA processes.

Let $\{X_t\} \sim ARMA(p,q)$

$$\phi(B) X_t = \theta(B) Z_t$$

$Z_t \sim WN(0, \sigma^2)$, where

$\phi(z) \neq 0$ (degree p)

$\theta(z) \neq 0$ (degree q).

when $|z| = 1$.

Then there exist polynomials

$$\phi^*(z), \quad \theta^*(z)$$

which are non zero for $|z| \leq 1$, of degree p and q respectively, such that $\{X_t\}$ has the causal and invertible representation

$$\phi^*(B) X_t = \theta^*(B) Z_t^*$$

where $\{Z_t^*\}$ is another white noise process.

Define

$$\phi^*(z) = \phi(z) \prod_{j=r+1}^{p} \left(\frac{1 - \xi_j z}{1 - \xi_j^{-1} z} \right)$$

$$\theta^*(z) = \theta(z) \prod_{j=s+1}^{q} \left(\frac{1 - \omega_j z}{1 - \omega_j^{-1} z} \right)$$

where $\xi_{r+1}, \xi_{r+2}, \ldots, \xi_p$

$\omega_{s+1}, \omega_{s+2}, \ldots, \omega_q$

are the reciprocal roots that lie within the unit circle.

Consider the process

$$Z_t^* = \phi^*(B)[\theta^*(B)]^{-1} X_t$$

$$= \left\{ \prod_{j=r+1}^{p} \left(\frac{1 - s_j B}{1 - s_j^{-1} B} \right) \right\} \left\{ \prod_{j=s+1}^{q} \left(\frac{1 - w_j B}{1 - w_j^{-1} B} \right)^{-1} \right\} Z_t .$$

For this process, it follows that

$$G_{Z^*}(z) = \sigma^2 \left\{ \prod_{j=r+1}^{p} |s_j|^2 \right\} \left\{ \prod_{j=s+1}^{q} |w_j|^2 \right\}$$

by the method above.

$\therefore \{Z_t^*\}$ is white noise

$\therefore \quad \phi^*(B) X_t = \theta^*(B) Z_t^*$

with

- $\{Z_t^*\} \sim WN(0, \sigma^{*2})$
- $\phi^*(z) \neq 0 \quad$ for all $|z| \leq 1$
- $\theta^*(z) \neq 0 \quad$ for all $z \leq 1$.

$\therefore \{X_t\}$ has a causal/invertible representation.

PACF FOR ARMA(p,q)

Recall the function α_x

$$\alpha_x(0) = 1$$

$$\alpha_x(h) = \phi_{hh} \quad \text{where} \quad h \geq 1$$

ϕ_{hh} is the last component of

$$\underset{\sim}{\phi}_h = \Gamma_h^{-1} \underset{\sim}{\gamma}_x(h)$$

$$[\Gamma_h]_{ij} = [\gamma_x(i-j)] \quad \underset{\sim}{\gamma}_x(h) = [\gamma_x(1), \ldots, \gamma_x(h)]^T$$

γ_x - acv sequence for $\{X_t\}$

We can compute the sample acf by replacing data-based quantities

i.e. $\hat{\alpha}_n(0) = 1$

$$\hat{\alpha}_n(h) = \hat{\phi}_{hh} \qquad h \geq 1$$

$$\hat{\underset{\sim}{\phi}}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_n(h)$$

- only numerical calculations of

$$\phi_{hh} \text{ and } \hat{\phi}_{hh}$$

available in general.

Recall

- For AR(p)

$$\alpha_x(h) = 0 \qquad h > p$$

- For MA(q)

$$\gamma_x(h) = 0 \qquad h > q.$$

- This helps identify the process from sample data.

i.e

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & 1 \le n < m \\ \sum_{j=1}^{p} \phi_j X_{n-j+1} + \sum_{j=1}^{q} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) & n \ge m \end{cases}$$

and

$$E\left[(X_{n+1} - \hat{X}_{n+1})^2 \right] = \sigma^2 r_n .$$

Then if

$$\hat{W}_{n+1} = P_n W_{n+1}$$

and

$$\hat{X}_{n+1} = P_n X_{n+1}$$

it follows that by linearity of P_n

$$\hat{W}_t = \sigma^{-1} \hat{X}_t \qquad t = 1, \ldots, m$$

$$\hat{W}_t = \frac{1}{\sigma}\left[\hat{X}_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}\right] \qquad t > m$$

$$\therefore \; X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t) \qquad t \geq 1.$$

Therefore we can substitute back in to the previous equations to get forecasts for $\{X_t\}$.

Then from the innovations algorithm

$$\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}) & 1 \le n < m \\ \sum_{j=1}^{q} \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}) & n \ge m \end{cases}$$

coefficients θ_{nj} $\forall n, j$

m.s.e $r_n = E[(W_{n+1} - \hat{W}_{n+1})^2]$

are computed by the algorithm.

Note that $\theta_{nj} = 0$ if $n \ge m$ and $j > q$.

Let $\kappa(i,j) = E[W_i W_j] \quad i,j \geq 1.$

It follows that

$$\kappa(i,j) = \begin{cases} \frac{1}{\sigma^2}\gamma_X(i-j) & 1 \leq i,j \leq m \\ \frac{1}{\sigma^2}\left[\gamma_X(i-j) - \sum_{r=1}^{p}\phi_r \gamma_X(r-|i-j|)\right] & \min(i,j) \leq m \leq \max(i,j) \leq 2m \\ \sum_{r=0}^{q}\theta_r \theta_{r+|i-j|} & \min(i,j) \geq m \\ 0 & \text{otherwise} \end{cases}$$

- found using careful calculation of acvs of $\{W_t\}$ using standard methods

Assume $\sigma_0 = 1$.

$\gamma_X(\cdot)$ can be computed using previous techniques.

For the innovations algorithm, we need to input autocovariances into prediction mechanism

- the trick is to forecast $\{W_t\}$ and transform to a forecast for $\{X_t\}$

3.1.3 Forecasting ARMA(p,q)

For the process $\{X_t\}$

$$\phi(B)X_t = \theta(B)Z_t$$

$Z_t \sim WN(0,\sigma^2)$, the innovations algorithm gives a suitable forecasting method.

Let $\{W_t\}$ be defined by

$$W_t = \begin{cases} \frac{1}{\sigma} X_t & t = 1, \ldots, m \\ \frac{1}{\sigma} \phi(B) X_t & t > m \end{cases}$$

where $m = \max(p,q)$.

We can use the usual asymptotic arguments to test for AR(p) / MA(q)

i.e.

$$\sqrt{n}\left(\hat{\alpha}_n(h) - \alpha_x(h)\right) \xrightarrow{\mathcal{L}} N(0,1)$$

i.e. $$\hat{\alpha}_n(h) \sim AN\left(0, \frac{1}{n}\right)$$

i.e we expect

$$|\hat{\alpha}_n| < \frac{1.96}{\sqrt{n}}$$

with approximately 95% prob.

3.2 Estimation and Model Selection

For general stationary processes, moment-based estimation of μ_x, $\gamma_x(h)$, σ^2 and $\rho_x(h)$ is possible.

For ARMA(p,q) processes, we need to estimate $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots \theta_q$. The previous moment-based methods can be implemented, but are generally **inefficient**

[i.e the estimators do not have the smallest possible variances]

For the ARMA(p,q) process

$$\phi(B) X_t = \theta(B) Z_t$$

with $\{Z_t\} \sim WN(0, \sigma^2)$, special algorithms exist to compute estimates

- YULE WALKER
- BURG.

Often, we can improve efficiency if LIKELIHOOD methods are available.

YULE-WALKER ALGORITHM FOR AR(p)

Suppose $\{X_t\} \sim AR(p)$

$$\phi(B)X_t = Z_t$$

i.e $$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t .$$

$$\therefore E[X_t^2] = \phi_1 E[X_{t-1}X_t] + \ldots + \phi_p E[X_{t-p}X_t]$$

$$= \phi_1 \gamma_X(1) + \ldots + \phi_p \gamma_X(p)$$

$$E[X_t X_{t-1}] = \phi_1 E[X_{t-1}X_{t-1}] + \ldots + \phi_p E[X_{t-p}X_{t-1}]$$

$$= \phi_1 \sigma_X^2 + \phi_2 \gamma_X(1) + \ldots + \phi_p \gamma(p-1)$$

etc.

$$\therefore \quad \Gamma_p \underset{\sim}{\phi} = \underset{\sim}{\gamma}_x(p)$$

YULE WALKER EQUATIONS

where

$$\Gamma_p = \left[\gamma_x(i-j)\right]_{i,j=1}^{p}$$

$$\underset{\sim}{\gamma}_x(p) = \left(\gamma_x(1), \ldots, \gamma_x(p)\right)^T$$

and

$$\sigma_w^2 = \gamma_x(0) - \underset{\sim}{\phi}^T \underset{\sim}{\gamma}_x(p)$$

Replacing the Theoretical acv matrix/vector by their sample counterparts yields an equation for

$$\underset{\sim}{\phi} = (\phi_1, \ldots, \phi_p)^T$$

that can be solved explicitly.

i.e

$$\hat{\underset{\sim}{\phi}} = \hat{\Gamma}_n^{-1} \hat{\underset{\sim}{\gamma}}_n(p)$$

$$= \hat{R}_n^{-1} \hat{\underset{\sim}{\rho}}_n(p)$$

where $\hat{\Gamma}_n$ is the matrix of sample acvs

$\hat{\underset{\sim}{\gamma}}_n(p)$ - sample acv vector

$$= (\hat{\gamma}_n(1), \ldots, \hat{\gamma}_n(p))^T.$$

$\hat{R}_n^{-1}$ is the (inverse) sample correlation matrix

$$\hat{\underset{\sim}{\rho}}_n(p) = (\hat{\rho}_n(1), \ldots, \hat{\rho}_n(p))^T$$

Also

$$\hat{\sigma}^2 = \hat{\gamma}_n(0)\left[1 - \hat{\underset{\sim}{\rho}}_n^T(p)\, \hat{R}_n^{-1}\, \hat{\underset{\sim}{\rho}}_n(p)\right]$$

Asymptotic Gaussianity

For large n

$$\hat{\underset{\sim}{\phi}}_n \sim AN\left(\underset{\sim}{\phi}, \frac{\sigma^2}{n}\Gamma_p^{-1}\right)$$

[This can give approximate likelihood inference]

NOTE It can be shown that the Yule-Walker estimates given by

$$\hat{\underset{\sim}{\phi}} = \hat{R}_n^{-1} \hat{\underset{\sim}{\rho}}_n(p)$$

are a causal solution

and thus the fitted covariance sequence

$$\{\hat{\gamma}_x(h)\} \qquad h = 0, 1, \ldots$$

derived from $\hat{\underset{\sim}{\phi}}$ satisfies

$$\hat{\gamma}_x(h) - \hat{\phi}_1 \hat{\gamma}_x(h-1) - \cdots - \hat{\phi}_p \hat{\gamma}_x(h-p)$$

$$= \begin{cases} 0 & h = 1, \ldots p \\ \hat{\sigma}^2 & h = 0. \end{cases}$$

Furthermore, using the same result

$$X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p} = Z_t$$

where $Z_t \sim WN(0, \hat{\sigma}^2)$, so the residuals from the fit should be an uncorrelated sequence with homoscedastic variance σ^2.

i.e

$$e_t = X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p}$$

should be an approximate white noise sequence.

NOTE We can construct an AR(p) process with arbitrary ACV γ_x by defining

$$\underset{\sim}{\phi} = \Gamma_{p+1}^{-1} \underset{\sim}{\gamma}_x(p)$$

where

$$\underset{\sim}{\gamma}_x(p) = (\gamma_x(0), \ldots, \gamma_x(p))^T$$

$$\Gamma_{p+1} = \left[\gamma_x(i-j) \right]_{i,j=1}^{p+1}.$$

YULE-WALKER FOR ARMA(p,q)

Solving the equivalent system for ARMA(p,q) is not straightforward.

If $\phi(B)X_t = \theta(B)Z_t$

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

Then the Yule-Walker equations become

$$\hat{\gamma}_n(k) - \hat{\phi}_n(k-1) - \dots \hat{\gamma}_n(k-p)$$

$$= \sigma^2 \sum_{j=k}^{q} \theta_j \psi_{j-k}$$

for $0 \le k \le p+q$

but this is nonlinear in $\underset{\sim}{\phi}, \underset{\sim}{\theta}$.

Note Fitted Yule-Walker AR(p) model is the model with the coefficients

$$\hat{\underset{\sim}{\phi}}$$

input to the model.

The parameters $\underset{\sim}{\phi} = (\phi_1, \ldots, \phi_p)^T$

are model order dependent, so we introduce a second index, and write

$$\underset{\sim}{\phi}_p = (\phi_{p1}, \ldots, \phi_{pp})^T$$

- the first index indicates the model order.

The coefficients

$$\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$$

are the coefficients in the best linear predictor of X_{p+1} in terms of $X_1, \ldots, X_p$, with the assumption

$$\gamma_X(j) = \hat{\gamma}_n(j) \quad \text{for } j = 1, \ldots, p$$

↖ sample acf.

i.e $\hat{\phi}_{pp}$ is the partial autocorrelation coefficient at lag p.

BURG'S ALGORITHM FOR THE PACFS

Define, for $t = i+1, \ldots, n$, $0 \le i < n$

$$U_i(t) = x_{n+1+i-t} - \hat{x}^{F}_{n+1+i-t}$$

$$V_i(t) = x_{n+1+i-t} - \hat{x}^{B}_{n+1+i-t}$$

where

$$\hat{x}^{F}_{n+1+i-t} \quad , \quad \hat{x}^{B}_{n+1+i-t}$$

are the best linear predictors given the preceding / following i observations respectively

F - forward

B - backward.

It follows that

$$\text{INIT}: U_0(t) = V_0(t) = x_{n+1-t}$$

$$\text{REC}: U_i(t) = U_{i-1}(t-1) - \phi_{ii} V_{i-1}(t)$$

$$V_i(t) = V_{i-1}(t) - \phi_{ii} U_{i-1}(t-1)$$

We choose ϕ_{ii} to minimize

$$\sigma_i^2 = \frac{1}{2(n-i)} \left\{ \sum_{t=i+1}^{n} [U_i^2(t) + V_i^2(t)] \right\}$$

The Burg recursions are simplified by performing the minimization analytically.

BURG'S ALGORITHM

INITIALIZE: $$d(1) = \sum_{t=2}^{n} \left(u_0^2(t-1) + v_0^2(t) \right)$$

RECURSION: Set

$$\hat{\phi}_{ii}^{B} = \frac{2}{d(i)} \left\{ \sum_{t=i+1}^{n} v_{i-1}(t)\, u_{i-1}(t-1) \right\}$$

$$d(i+1) = \left(1 - (\hat{\phi}_{ii}^{B})^2\right) d(i) - v_i^2(i+1) - u_i^2(n)$$

$$\sigma_i^2 = \frac{\left(1 - (\hat{\phi}_{ii}^{B})^2\right) d(i)}{2(n-i)}$$

for $i = 1, \ldots, p$.

NOTE Burg's Algorithm, in conjunction with the Levinson-Durbin algorithm, allows calculation of

$$\hat{\phi}_{11}^{B}, \ldots \hat{\phi}_{pp}^{B}$$

which are estimates of the coefficients in the AR(p) model.

- we again replace

$$\gamma_X(h)$$

by

$$\hat{\gamma}_m(h)$$

in the Levinson-Durbin algorithm.

Also, $\underset{\sim}{\hat{\phi}}_p^B \sim AN(\underset{\sim}{\phi}, \frac{\sigma^2}{n}\Gamma_p^{-1})$

Innovations Algorithm For MA(q)

We can use the innovations algorithm to construct the estimates

$$(\hat{\theta}_{q1}, \ldots, \hat{\theta}_{qq})^T$$

of $\underset{\sim}{\theta}$ in the MA(q) model.

Again, we replace

$$\gamma_X(h)$$

by

$$\hat{\gamma}_n(h)$$

in the algorithm.

Note These estimates are <u>not always consistent</u>

3.2.2 MAXIMUM LIKELIHOOD ESTIMATION

ML estmation typically produces the best possible estimates

- consistent
- efficient
- good mall sample properties.

Unfortunately it is hard to carry out MLE for non Gaussian time series.

Some approximation methods can be used.

Let $\{X_t\}$ be a Gaussian time series for which the acvf is $K(i,j)$
- we assume $\{X_t\}$ is zero mean, so

$$K(i,j) = E[X_i X_j]$$

Let $\underset{\sim}{X}_n = (X_1, \ldots, X_n)^T$

$\underset{\sim}{\hat{X}}_n = (\hat{X}_1, \ldots, \hat{X}_n)^T$

where $\hat{X}_1 = 0$

$$\hat{X}_j = E[X_j | X_1, \ldots, X_{j-1}] \qquad j \geq 2$$

$$= P_j X_{j-1}$$

- always available from standard properties of MVNormal dist^n.

Let

$$\Gamma_n = E[\underset{\sim}{X}_n \underset{\sim}{X}_n^T]$$

Then

$$L(\Gamma_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Gamma_n|^{1/2}}$$

$$\exp\left\{-\frac{1}{2} \underset{\sim}{X}_n^T \Gamma_n^{-1} \underset{\sim}{X}_n\right\}$$

If n is large, computation of

$|\Gamma_n|$

Γ_n^{-1}

may be computationally prohibitive.

However, the innovations algorithm can be used to avoid direct calculation of the determinant and inverse.

The innovations algorithm returns

$$\theta_{ij} \quad j = 1, \ldots, i \quad i = 1, 2, \ldots$$

as coefficients.

Let $C_n = \begin{bmatrix} 1 & & & & \\ \theta_{11} & 1 & & & \\ \theta_{22} & \theta_{21} & 1 & & \\ \theta_{33} & \theta_{32} & \theta_{31} & \ddots & \\ \vdots & \vdots & \vdots & \ddots & \\ \theta_{n-1,n-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$

Then, as before

$$\underset{\sim}{X}_n = C_n(\underset{\sim}{X}_n - \underset{\sim}{\hat{X}}_n)$$

The elements of $(\underset{\sim}{X}_n - \underset{\sim}{\hat{X}}_n)$ are uncorrelated, with covariance matrix

$$D_n = \text{diag}(V_0, V_1, \ldots, V_{n-1})$$

where

$$V_{j-1}, \quad j = 1, \ldots, n-1$$

are the prediction variances computed in the innovations algorithm.

Thus

$$\Gamma_n = C_n D_n C_n^T$$

$$\underset{\sim}{x}_n^T \Gamma_n^{-1} \underset{\sim}{x}_n = (\underset{\sim}{x}_n - \hat{\underset{\sim}{x}}_n)^T D_n^{-1} (\underset{\sim}{x}_n - \hat{\underset{\sim}{x}}_n)$$

$$= \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{V_{j-1}}$$

and

$$|\Gamma_n| = V_0 V_1 V_2 \ldots V_{n-1}.$$

i.e the innovations algorithm gives a means of diagonalizing the covariance matrix.

1.e

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n V_0 V_1 \ldots V_{n-1}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{V_{j-1}}\right\}$$

Typically, Γ_n will be parameterized (i.e. if $\{X_t\} \sim ARMA(p,q)$), so we will be estimating a finite set of parameters

The asymptotic properties of ML estimates are well known.

For ARMA processes

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \le n < m \\ \sum_{j=1}^{p} \phi_j X_{n+1-j} + \sum_{j=1}^{q} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & n \ge m \end{cases}$$

where $m = \max(p, q)$

$$V_n = E\left[(X_{n+1} - \hat{X}_{n+1})^2\right] = \sigma^2 r_n$$

where $\{\theta_{nj}\}$ and $\{r_n\}$ are found using the innovations algorithm.

NON-GAUSSIAN SERIES

This likelihood approach can be used for approximate ML estimation even when $\{X_t\}$ is non-Gaussian.

"GAUSSIAN ESTIMATION"

- we merely need the form of $\kappa(i,j)$ to compute Γ_n and give the entries in the innovations algorithm.

The asymptotic properties of the resulting estimates are still good.

ARMA(p,q) likelihood is

$$L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2)$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} \left\{ \prod_{j=1}^{n} r_{j-1} \right\}^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) \right\}$$

$R(\underset{\sim}{\phi}, \underset{\sim}{\theta})$

where

$$S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j(\underset{\sim}{\phi}, \underset{\sim}{\theta}))^2}{r_{j-1}}$$

- sum of standardized prediction errors squared.

This likelihood allows estimation of $(\underset{\sim}{\phi}, \underset{\sim}{\theta})$ and σ^2 to be separated

$$(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}}) = \operatorname{argmax} S(\underset{\sim}{\phi}, \underset{\sim}{\theta}) / R(\underset{\sim}{\phi}, \underset{\sim}{\theta})$$

$$\equiv \operatorname{argmax} L(\underset{\sim}{\phi}, \underset{\sim}{\theta})$$

where

$$L(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \log\left(\frac{1}{n} S(\underset{\sim}{\phi}, \underset{\sim}{\theta})\right) + \frac{1}{n} \sum_{j=1}^{n} \log r_{j-1}$$

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}})$$

Note There is no guarantee that the resulting estimates will be such that the estimated process is CAUSAL and INVERTIBLE.

- we can impose these constraints when performing the maximization

By reparameterizing we may be able to impose the constraints directly

- a natural approach would be to parameterize using the RECIPROCAL ROOTS

$\xi_1, \ldots, \xi_p$ and $\omega_1, \ldots, \omega_q$.

LEAST SQUARES Could choose $\underset{\sim}{\phi}, \underset{\sim}{\Theta}$ to minimize the prediction errors squared

i.e.

$$(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\Theta}}) = \arg\min S(\underset{\sim}{\phi}, \underset{\sim}{\Theta})$$

$$\tilde{\sigma}^2 = \frac{S(\tilde{\underset{\sim}{\phi}}, \tilde{\underset{\sim}{\Theta}})}{n-p-q}$$

- another numerical operation; can introduce constraints etc.

Thus to fit the models we have the options to use

MLE	LIKELIHOOD
LS	LEAST SQUARES

For Pure AR, we have

MLE	LIKELIHOOD
LS	LEAST SQUARES
YW	MOMENTS
BURG.	MOMENTS

All procedures produce CONSISTENT estimates Likelihood methods theoretically preferable in terms of efficiency.

However, for pure AR the four methods seem to perform equally well even for $n = 100$ in simulation studies.

Differences are likely to be more evident for ARMA.

Standard Errors + Confidence Intervals

By usual ML Theory, if $\underset{\sim}{\beta} = (\underset{\sim}{\phi}, \underset{\sim}{\theta})^T$, then asymptotically

$$\hat{\underset{\sim}{\beta}} \sim AN\left(\underset{\sim}{\beta}, \frac{1}{n} V(\underset{\sim}{\beta})\right)$$

where $V(\beta)$ is the inverse Fisher Information.

i.e.

$$\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{d} N(0, V(\underset{\sim}{\beta}))$$

where we can estimate the asymptotic variance $V(\underset{\sim}{\beta})$ by using the <u>observed</u> information

$$\psi(\underset{\sim}{\beta}) = \frac{1}{n}\left[-\frac{\partial^2 \log L(\underset{\sim}{\beta})}{\partial \beta_i \, \partial \beta_j}\right]_{i,j=1}^{p+q}$$

i.e we replace $V(\underset{\sim}{\beta})$ by

$$[\psi(\hat{\underset{\sim}{\beta}})]^{-1}$$

3.2.3 Model Selection

We can fit a series of models to any given data set: to choose the "best" model we can consider different issues:

- Fidelity to the data
- Model complexity
- Residuals
- Forecasting performance.

AKAIKE INFORMATION CRITERION (AIC)

The AIC model selection criterion is used widely in statistics: it trades off goodness of fit (fidelity to the data) with model complexity (number of parameters fitted).

For a model M with parameters $\underset{\sim}{\beta}$ yielding likelihood $L_M(\underset{\sim}{\beta})$

The AIC is defined as

$$AIC(M, \hat{\underset{\sim}{\beta}}) = -2 \log L_M(\hat{\underset{\sim}{\beta}}) + 2d$$

where d is the dimension of $\underset{\sim}{\beta}$.

For the ARMA(p,q) model

$$\underset{\sim}{\beta} = (\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2) \quad \therefore \text{ dimension } p+q+1$$

so the AIC is

$$AIC(M_{p,q}, \hat{\underset{\sim}{\beta}})$$

$$= -2 \ell_m(\hat{\underset{\sim}{\beta}}, \frac{S(\hat{\underset{\sim}{\beta}})}{n}) + 2(p+q+1)$$

$\ell_m = \log L_{M_{p,q}}$; $\hat{\sigma}^2$

This is easily computed (using R) once the model has been fitted.

We fit models of different orders and choose the model for which AIC is smallest : usually

"smallest" - "most negative"

The AIC is a "within sample" criterion

- it looks at fit to data already observed.

THIS MIGHT NOT BE THE OBJECTIVE OF THE ANALYSIS!

- WE MIGHT WISH TO PERFORM FORECASTING.

Opinions on the AIC seem to be mixed....

- it is widely used in time series and other fields

- however, it is also widely known to have (serious) flaws.

SMALL SAMPLE : Does not penalize for complexity to a large enough degree.
⇒ chooses models that are to complex

LARGE SAMPLE : INCONSISTENT!

i.e does not pick the true model wp 1 as $n \to \infty$

To fix the large sample problem :

Schwarz proposed the BIC

"BAYESIAN INFORMATION CRITERION"

In general, the BIC takes the form

$$BIC(M, \hat{\beta}) = -2L_M(\hat{\beta}, \frac{s(\hat{\beta})}{n}) + d\log n.$$

where n is the sample size.

i.e. the penalty $2d$ (AIC) is replaced by $d\log n$ (BIC).

To fix the small sample behaviour:

Hurvich + Tsai proposed the AICC

$$AICC(M_{p,q}, \underset{\sim}{\hat{\beta}})$$

$$= -2\mathcal{L}(\underset{\sim}{\hat{\beta}}, \hat{S}(\underset{\sim}{\hat{\beta}})) + \frac{2(p+q+1)n}{n-p-q-2}$$

$L = \log L_{n,\sigma}$.

- again easily computed after $\underset{\sim}{\hat{\beta}}$ computed.

- note that $\frac{n}{n-p-q-2} \longrightarrow 1$ as $n \to \infty$

so

AICC ad AIC are asymptotically equivalent.

AICC is a bias-corrected version of the AIC obtained by making a bias correction in the estimation of $\hat{\sigma}^2$.

ML : $$\hat{\sigma}^2 = \frac{S(\hat{\beta})}{n}$$

UNBIASED : $$\hat{\sigma}^2 = \frac{S(\hat{\beta})}{n-p-q-2}.$$

However, this process of correction does <u>not</u> correct the large sample problem

AICC is <u>still not consistent</u>!

As $n \to \infty$, $\log n \gg 2$

$\therefore$ The penalty per parameter is greater under the BIC

$\therefore$ BIC penalizes complexity to a greater degree.

As a result, it transpires that BIC IS a consistent model selection criterion in time series and elsewhere.

Good small sample performance is not guaranteed, but will be reasonably similar to AIC.

[$n \geq 8 \Leftrightarrow \log n \approx 2$].

NOTE For non-likelihood approaches (LS, YW, BURG) we can still use AIC/AICC/BIC in an approximate likelihood setting: we obtain parameter estimates $\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}}, \hat{\sigma}^2$, then

plug these into the innovations algorithm to obtain $\hat{X}_j, r_{j-1}$ for $j = 1, \ldots, n$, then use the **GAUSSIAN** Likelihood

$$L(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \sigma^2) = \frac{1}{(2\pi)^{n/2}} \left(\frac{1}{\sigma^2}\right)^{n/2} \left\{\prod_{j=1}^{n} r_{j-1}\right\}^{-1/2}$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right\}.$$

3.2.4 Residual Checking

For ARMA(p,q) models, the residual sequence $\{\hat{W}_t\}$ given by

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}})}{\{r_{t-1}(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}})\}^{1/2}}$$

i.e the one step ahead prediction errors standardized by $\hat{r}_{t-1}^{1/2} = \{r_{t-1}(\hat{\underset{\sim}{\phi}}, \hat{\theta})\}^{1/2}$

Note that $\{\hat{W}_t\}$ estimates

$$W_t(\underset{\sim}{\phi}, \underset{\sim}{\theta}) = \frac{X_t - \hat{X}_t(\underset{\sim}{\phi}, \underset{\sim}{\theta})}{\{r_{t-1}(\underset{\sim}{\phi}, \underset{\sim}{\theta})\}^{1/2}}$$

and $W_t \xrightarrow{\text{m.s.}} Z_t$ as $t \to \infty$.

$\therefore$ $\{\hat{W}_t\}$ should be approximately white noise with variance σ^2 if the model selected is correct.

We can check $\{\hat{W}_t\}$ is

- uncorrelated / independent
- normally distributed

We can also inspect $\{\hat{R}_t\}$

$$\hat{R}_t = \frac{\hat{W}_t}{\hat{\sigma}}$$

where $\hat{\sigma}$ is an estimate of the residual variance. We check $\{\hat{R}_t\} \dot{\sim} WN(0,1)$.

4 Non-stationary and Seasonal models

Models with structured nonstationary components can be constructed as straightforward extensions to ARMA models.

- Integrated Models
- Seasonal Models

We use the techniques introduced earlier to extend ARMA to

ARIMA

Seasonal Arima. (SARIMA)

NOTE

We can obtain non-stationary processes by introducing variation in

mean μ_t

variance σ_t

through time

Often, such non-stationarity can be captured by

STATE-SPACE MODELS

6.1 ARIMA MODELS

DEFINITION

If d is a non-negative integer, then

$\{X_t\}$ is an ARIMA(p,d,q) process if $\{X_t\}$ defined by

$$Y_t = (1-B)^d X_t$$

is a causal ARMA(p,q) process.

i.e $\{X_t\}$ satisfies

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

where $\{Z_t\} \sim WN(0,\sigma^2)$, with the reciprocal roots of $\phi(z) = 0$ inside the unit circle.

$\{X_t\}$ is stationary iff $d=0$;
if $d \geq 1$ we can add a polynomial $(d-1)$
trend to any solution of the
ARIMA difference equation, and the
resulting process still satisfies the
equation.

Note that

$$\phi^*(z) = (1-z)^d \phi(z)$$

so $\phi^*(z)$ has a root of order d at
$z = 1$; this is the only root at $|z| = 1$.

Note If $X_t \sim ARIMA(p,d,q)$ then

$$Y_t = (1-B)^d X_t$$

gives that

$$Y_t \sim ARMA(p,q)$$

- so estimation of $(\underset{\sim}{\theta}, \underset{\sim}{\phi})$ and σ^2 proceeds via the $\{y_t\}$ series obtained by differencing the $\{x_t\}$ series d times

$$y_t = (1-B)^d x_t$$

EXAMPLE Compare

STATIONARY ARMA(p,q) ≡ AR(2)

$$(1-0.6B)(1-0.95B)X_t = Z_t$$

NON STATIONARY ARIMA(p,q) ≡ ARIMA(1,1,0)

$$(1-0.6B)(1-B)X_t = Z_t$$

where $Z_t \sim IID(0,1)$

(in simulation, $Z_t \sim N(0,1)$)

- can be difficult to distinguish the two cases using the usual sample statistics.

However, as a general rule, if we observe a time series with a slowly decaying ACF, we try ARIMA models with $d = 1, 2, \ldots$

i.e we difference the data once, twice etc, and replot the sample ACF/PACF.

- for some d, we may find a slowly decaying ACF.

However, it is possible to construct nonstationary models with a <u>different</u> form of slowly decaying sample ACF.

Recall that for ARIMA(p,d,q)

$$\phi^*(B)X_t = (1-B)^d \phi(B) X_t = \theta(B) Z_t$$

where

$$\phi^*(z) = 0$$

has a root $z = 1$ (or $z \to 1$ for non-stationary looking stationary processes).

Consider a model where

$$\phi^*(z) = 0$$

has a root near/at $z = e^{i\omega}$

for some $\omega \in (-\pi, \pi)$

Here, if $z = e^{i\omega}$ is a root, then

$$|z| = |e^{i\omega}| = 1$$

so the model is <u>nonstationary</u>

Consider the reciprocal root

$$\xi = \frac{1}{r} e^{-i\omega}$$

and the model based on

$$\begin{aligned}\phi^*(B) &= (1 - \xi B)(1 - \bar{\xi} B) \\ &= 1 - 2\mathrm{Re}(\xi) B + |\xi|^2 B^2 \\ &= 1 - \left(\frac{2}{r}\cos\omega\right) B + \left(\frac{1}{r^2}\right) B^2\end{aligned}$$

It follows that, for this AR(2) process

$$\rho_X(h) = \left(\frac{1}{r}\right)^h \frac{\sin(h\omega+\lambda)}{\sin(\lambda)} \qquad h=0,1,2,\ldots$$

where

$$\lambda = \arctan\left[\left(\frac{r^2+1}{r^2-1}\right)\tan(\omega)\right]$$

Using eg MAPLE, we find that

$$\rho_X(h) \longrightarrow \cos(h\omega)$$

as $r \longrightarrow 1^+$

(i.e as the model tends to a non-stationary model)

i.e. we have a slowly decaying (due to $(\frac{1}{r})^h$ term) sinusoidal autocorrelation function.

- the ACF is not monotonic decreasing (as in the previous almost non-stationary case)

This process exhibits

SEASONAL PERSISTENCE

with seasonality determined by ω.

6.2 Unit Roots

The presence of a unit root of the AR polynomial equation

$$\phi(z) = 0$$

i.e a solution where $|z| = 1$ renders the resulting AR/ARMA process nonstationary

Therefore it is crucial to be able to identify processes where a unit root is present, and to distinguish this from a root near to but not equal to one in modulus.

THE DICKEY-FULLER TEST FOR UNIT ROOT

Suppose $\{X_t\}$ satisfies

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

where $Z_t \sim WN(0, \sigma^2)$ so that $E[X_t] = \mu$. with $|\phi| < 1$.
Under this model

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, 1 - \phi^2)$$

as $n \to \infty$, so we have a Z-test type result.

If $|\phi| = 1$, this asymptotic result does not hold.

Suppose $\phi = 1$. Then

$$(1-B)X_t = X_t - X_{t-1} = \mu(1-\phi) + (\phi-1)X_t + Z_t$$

$$= \phi_0^* + \phi_1^* X_t + Z_t \qquad (*)$$

where $\phi_0^* = \mu(1-\phi)$

$\phi_1^* = \phi - 1$.

As $\{Z_t\} \sim WN(0, \sigma^2)$, we can estimate ϕ_0^* and ϕ_1^* via least-squares in equation $(*)$

$\hat{\phi}_1^*$ estimated using LS has estimated standard error

$$\text{e.s.e.}\left(\hat{\phi}_1^*\right) = \frac{S}{\sqrt{\sum_{t=2}^{n}(x_{t-1}-\bar{x})^2}}$$

where

$$S = \frac{\sum_{t=2}^{n}\left(x_t - x_{t-1} - \hat{\phi}_0^* - \hat{\phi}_1^* x_{t-1}\right)^2}{n-3}$$

by the usual results for LS.

A Wald-type test statistic is

$$t = \frac{\hat{\phi}_1^*}{e.s.e(\hat{\phi}_1^*)}$$

As $n \to \infty$, t has a limiting distribution computed by Dickey and Fuller under the hypotheses

$$H_0 : \phi_1^* = 0$$

that is

$$H_0 : \phi_1 = 1.$$

One sided alternative $H_a : \phi_1 < 1$.

The technique can be extended to AR(p)

If

$$X_t - \mu = \sum_{i=1}^{p} \phi_i (X_{t-i} - \mu) + Z_t$$

then

$$(1-B)X_t = \phi_0^* + \phi_1^* X_{t-1} + \sum_{i=2}^{p} \phi_i^* (1-B) X_{t-i+1} + Z_t .$$

where

$$\phi_0^* = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$$

$$\phi_1^* = \sum_{i=1}^{p} \phi_i - 1$$

$$\phi_j^* = -\sum_{i=j}^{p} \phi_i \qquad j = 2, \ldots p .$$

$X_t \sim AR(p)$ with unit root

$$\Rightarrow (1-B)X_t \sim AR(p-1)$$

$\therefore$ Again a unit root for $\{X_t\}$ can be tested by assessing

$$\phi_1^* = 0$$

We estimate ϕ_1^* by LS on the regression model of $(1-B)X_t$ on

$$1, X_{t-1}, (1-B)X_{t-1}, \ldots, (1-B)X_{t-p+1}$$

The resulting test statistic

$$t = \frac{\hat{\phi}_1^*}{\text{e.s.e}(\hat{\phi}_1^*)}$$

has the Dickey-Fuller distribution adjusted

UNIT ROOTS IN MA PROCESSES

If $\{X_t\} \sim ARMA(p,q)$ is causal and invertible, $\{X_t\}$ defined by

$$Y_t = (1-B)X_t$$

is $ARMA(p,q+1)$ as, if $\{Z_t\} \sim WN(0,\sigma^2)$

$$\phi(B)X_t = \theta(B)Z_t$$

Then

$$Y_t = (1-B)X_t$$

$$\Rightarrow \phi(B)Y_t = (1-B)\theta(B)Z_t$$

$\therefore$ The MA polynomial for Y_t is

$$\theta_Y(z) = (1-z)\theta(z).$$

Thus a test for a unit root in the MA polynomial for Y_t is equivelent to a test for "over differencing" of a stationary/causal/invertible ARMA process.

The test is achieved using Likelihood-Ratio tests, based on Gaussian likelihoods or approximations.

The asymptotic distribution of the LR test statistic is non-standard, but has been computed.

4.3 Forecasting under non-stationarity.

Under general non-stationarity, forecasting is not readily achievable.

Recall that forecasting depended on knowledge of the autocovariance function in the stationary case. For the non stationary case, second moment function $K(i,j)$ - fed into innovations algorithm only if the second moment is finite for all t.

FORECASTING ARIMA

For the general ARIMA(p,d,q) model

$$(1-B)^d \phi(B) X_t = \theta(B) Z_t .$$

for $d \geq 1$ it is not straightforward to determine

$$E[X_t] \, , \, E[X_t X_{t+h}]$$

and thus the usual construction of best linear predictors cannot be used without further assumptions.

Let

$$(1-B)^d X_t = Y_t$$

where $\{Y_t\} \sim ARMA(p,q)$, and $\{Y_t\}$ is causal.

Suppose that $(X_{1-d}, \ldots, X_{-1}, X_0)$ is uncorrelated with Y_t.

Then

$$X_t = Y_t - \sum_{j=1}^{d} \binom{d}{j} (-1)^j X_{t-j}$$

by definition, for $t > 0$.

We wish to predict X_{n+h}.

By linearity of the prediction operator P_n

$$P_n X_{n+h} = P_n Y_{n+h} - \sum_{j=1}^{d} \binom{d}{j} (-1)^j P_n X_{n+h-j}$$

$P_n Y_{n+h}$ is computed using the methods described previously for ARMA(p,q).
- Y_t is uncorrelated with $(X_{-d}, \ldots, X_{-1}, X_0)$

$P_n X_{n+h-j}$ is computed recursively for $h = 1, 2, \ldots$

noting that

$$P_n X_{n+1-j} = X_{n+1-j}$$

for $j \geq 1$.

4.4 Seasonal ARMA Models

Recall that the lag-s difference operator ∇^s

$$\nabla^s X_t = X_t - X_{t-s}$$

could also be used to construct time series models. For example

ARIMA: $$(1-B)^d \phi(B) X_t = \theta(B) Z_t$$

Seasonal ARIMA: $$(1-B^s) \phi(B) X_t = \theta(B) Z_t.$$

This basic operator can be used to define more complex models.

Definition Seasonal ARIMA Model

If d, D are non-negative integers, then $\{X_t\}$ is a seasonal ARIMA process

$$X_t \sim \text{ARIMA}(p,d,q) \times (P,D,Q)_s$$

with period s if

$$Y_t = (1-B)^d (1-B^s)^D X_t$$

is a causal ARMA process determined by

$$\phi(B)\,\Phi(B^s)\,Y_t = \theta(B)\,\Theta(B^s)\,Z_t$$

where $\{Z_t\} \sim WN(0,\sigma^2)$

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad \text{order } p$$

$$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_P z^P \quad \text{order } P$$

$$\theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q \quad \text{order } q$$

$$\Theta(z) = 1 - \Theta_1 z - \dots - \Theta_Q z^Q \quad \text{order } Q$$

A SARIMA model allows for random variation in the seasonal patterns of seasonal data

- useful as, often, seasonal patterns are not constant.

Note : The differenced/fractional differenced series Y_t

$$Y_t = (1-B)^d (1-B^s)^D X_t$$

is also an ARMA process of the form

$$\phi^*(B) Y_t = \theta^*(B) Z_t$$

where

$$\phi^*(B) = \phi(B)\Phi(B^s) \quad \text{order } p+sP$$

$$\theta^*(B) = \theta(B)\Theta(B^s) \quad \text{order } q+sQ$$

We can estimate the coefficients using standard techniques.

In practice, we can try different plausible values of d, D, s, then compute AIC/BIC etc for the fitted model to select the optimal orders.

Normally we would try

$$d, D = 0, 1, 2, 3$$

$$s = 1, 4, 12 \quad \text{(monthly data)}$$
$$= 1, 7, 28 \quad \text{(daily data)}$$

Conditional on a given d, D, s, p, q, P, Q selection, estimation of ARMA parameters proceeds as usual using ML, with the same asymptotic behaviour etc.

4.5 NON-STATIONARITY DUE TO TREND

Previous non-stationary models address only issues of constant mean processes.

More generally, we may have non-constant mean processes. We can address estimation of trend / covariance structures simultaneously in an iterative procedure

- initialize
- given covariance structure, fit trend
- given fitted trend, estimate covariance structure
- iterate.

Suppose

$$Y_t = \underset{\sim}{x}_t^T \underset{\sim}{\beta} + W_t$$

where $\underset{\sim}{x}_t$ is a vector of covariates

$\underset{\sim}{\beta}$ is a fixed effects parameter

W_t is ARMA(p, q), zero mean.

$W_t \sim$ ARMA$(p, q)(\phi, \Theta)$

We wish to estimate

$$\underset{\sim}{\beta}$$

$$\underset{\sim}{\phi} = (\phi_1, \ldots, \phi_p)^T$$

$$\underset{\sim}{\Theta} = (\Theta_1, \ldots, \Theta_q)^T$$

Let

$$\mathrm{Cov}[\underset{\sim}{W}] = E[\underset{\sim}{W}\underset{\sim}{W}^T]$$

$$= \Gamma_n \qquad n \times n \text{ matrix.}$$

Γ_n is a function of (ϕ, θ). To estimate $\underset{\sim}{\beta}$ when

$$\Gamma_n = \sigma^2 I_n$$

we use ordinary least squares (OLS)

$$\hat{\underset{\sim}{\beta}} = (X^TX)^{-1}X^T\underset{\sim}{Y}$$

$$\mathrm{Var}[\hat{\underset{\sim}{\beta}}] = \sigma^2 (X^TX)^{-1}$$

$$\hat{\underset{\sim}{Y}} = X\hat{\underset{\sim}{\beta}} = X(X^TX)^{-1}X^T\underset{\sim}{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\underset{\sim}{Y} - \hat{\underset{\sim}{Y}})^T(\underset{\sim}{Y} - \hat{\underset{\sim}{Y}})$$

$$X = \begin{bmatrix} \underset{\sim}{x}_1^T \\ \\ \underset{\sim}{x}_n^T \end{bmatrix}$$ - DESIGN MATRIX

Usually X^TX is invertible, but if not $\hat{\underset{\sim}{\beta}}$ still satisfies

$$(X^TX)\hat{\underset{\sim}{\beta}} = X^T\underset{\sim}{Y}$$

but not uniquely.

These results hold in the Gaussian and non-Gaussian cases.

If Γ_n is not diagonal we use Generalized Least Squares (GLS)

$$\hat{\beta} = (X^T \Gamma_n^{-1} X)^{-1} X^T \Gamma_n^{-1} Y$$

$$\text{Var}[\hat{\beta}] = (X^T \Gamma_n^{-1} X)^{-1}$$

Let $V = \frac{1}{\sigma^2} \Gamma_n$

and let $V^{1/2}$ be the matrix square root of V, such that

$$V^{1/2} V^{1/2} = V.$$

Denote by $V^{-1/2}$ the inverse of $V^{1/2}$.

Then

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{W}$$

$$\Rightarrow V^{-1/2}\underset{\sim}{Y} = V^{-1/2}X\underset{\sim}{\beta} + V^{-1/2}\underset{\sim}{W}$$

Now

$$\text{Cov}[V^{-1/2}\underset{\sim}{W}] = V^{-1/2}\Gamma_n V^{-1/2}$$

$$= \sigma^2 I_n$$

$$\therefore \underset{\sim}{Y}^* = X^*\underset{\sim}{\beta} + \underset{\sim}{U}^*$$

$$\underset{\sim}{Y}^* = V^{-1/2}\underset{\sim}{Y}$$

$$X^* = V^{-1/2}X$$

$$\underset{\sim}{U}^* \sim WN(0, \sigma^2).$$

i.e given V, we can convert the problem to OLS.

We adopt the iterative scheme

① Estimate $\underset{\sim}{\beta}$, given V using OLS

② Estimate V, given $\underset{\sim}{\beta}$ using ML

to produce sequences of estimators

$$\hat{\underset{\sim}{\beta}}^{(0)}, \hat{\underset{\sim}{\beta}}^{(1)}, \ldots, \hat{\underset{\sim}{\beta}}^{(n)} \longrightarrow \hat{\underset{\sim}{\beta}} \quad \text{as } n \to \infty$$

$$V^{(0)}, V^{(1)}, \ldots V^{(n)} \longrightarrow V \quad \text{as } n \to \infty.$$

5. State Space Models

By "coupling" two stationary processes, we can construct more complex correlation patterns, and also construct processes that are non-stationary

(The first stationary process is functionally dependent on the second, so that we have dual (conditional and marginal) stationarity for the two processes, but marginal non-stationarity for the first.)

- Process 1, conditional on process 2, is a stationary process used to model the observed time series.

- Process 2 is used to model <u>latent structure</u>

  i.e structure that is not directly observable but whose influence is responsible for some of the variation observed in the data

We can make inference for parameters that govern the evolution of <u>each</u> process from a single time series.

Such dual models are useful in the modelling of

- financial time series
- missile-tracking systems
- EEG series

etc

but are also useful as representations of common stationary processes.

- ARMA
- ARMA

processes have "state-space" representations.

These representations can be computationally useful.

VECTOR-VALUED TIME SERIES

A vector-valued time series $\{\underset{\sim}{X}_t\}$

is a $(k \times 1)$ vector process that is potentially

- autocorrelated
- cross correlated

i.e if $\underset{\sim}{X}_t = (X_{t1}, \ldots, X_{tk})^T$

then

$$\gamma_{ij}(t,s) = E\left[(X_{ti} - \mu_{ti})(X_{sj} - \mu_{sj})\right]$$

is potentially non-zero, for all i, j, t, s.

We will mainly deal with stationary vector-valued or multivariate time series: for weak stationarity, we require

$$\underset{\sim}{\mu} = E[\underset{\sim}{X}_t] \qquad k \times 1$$

$$\Gamma(h) = \mathrm{Cov}[\underset{\sim}{X}_{t+h}, \underset{\sim}{X}_t] \qquad k \times k$$

both be independent of t.

i.e $$\Gamma_{\underset{\sim}{X}}(h) = [\gamma_{ij}(h)]_{i,j=1}^{k}$$

Here

$$\gamma_{ii}(h) \equiv \gamma_{X_i}(h)$$

$$\gamma_{ij}(h) = \gamma_{ji}(-h)$$

where

$$\gamma_{ij}(h) = \text{Cov}\left[X_{t+h,i}\,,\, X_{tj}\right]$$

Define also

$$R(h) = \left[\rho_{ij}(h)\right]_{i,j=1}^{k}$$

where

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\,\gamma_{jj}(0)}}$$

Many of the results and representations used in univariate time series also carry through to the multivariate case.

The equivalent of a white noise process in the univariate setting is a <u>multivariate</u> white noise process

$$\{\underset{\sim}{W}_t\} \sim WN(\underset{\sim}{0}, \{R_t\})$$

where

$$\{R_t\}$$

is a deterministic sequence of $k \times k$ matrices so that $E[\underset{\sim}{W}_t] = \underset{\sim}{0}$

$$E[\underset{\sim}{W}_t \underset{\sim}{W}_s^T] = \begin{cases} R_t & \text{if } t=s \\ 0 & \text{otherwise} \end{cases}$$

5.1 STATE-SPACE FORMULATION

A state-space model for multivariate time series $\{\underset{\sim}{Y}_t\}$ $(w \times 1)$ is specified by two equations:

OBSERVATION EQUATION:

$$\underset{\sim}{Y}_t = G_t \underset{\sim}{X}_t + \underset{\sim}{W}_t \qquad t = 0, 1, \ldots$$

where

$\{G_t\}$ is a sequence of $w \times v$ matrices (deterministic)

$$\{\underset{\sim}{W}_t\} \sim WN(\underset{\sim}{0}, \{R_t\})$$

i.e. $\underset{\sim}{Y}_t$ is a linear combination of elements of $\underset{\sim}{X}_t$

$\{\underset{\sim}{X}_t\}$ is a $(v \times 1)$ time series process following the **STATE EQUATION**

$$\underset{\sim}{X}_{t+1} = F_t \underset{\sim}{X}_t + \underset{\sim}{V}_t$$

where

$\{F_t\}$ is a sequence of $(v \times v)$ matrices (deterministic)

$$\{\underset{\sim}{V}_t\} \sim WN(\underset{\sim}{0}, \underset{\sim}{Q}_t)$$

and

$$E[\underset{\sim}{W}_t \underset{\sim}{V}_s^T] = 0 \quad \forall s,t.$$

To complete the model $\underset{\sim}{X}_1$ is uncorrelated with $\{\underset{\sim}{W}_t\}$, $\{\underset{\sim}{V}_t\}$.

NOTE It is possible to extend this model further by

(i) allowing $\{\underset{\sim}{W}_t\}$, $\{\underset{\sim}{V}_t\}$ to be correlated.

(ii) introducing slightly more sophisticated state equations

(iii) moving to non-linear state and/or observation equations

eg

$$\underset{\sim}{Y}_t = G_t \underset{\sim}{X}_t + \underset{\sim}{d}_t + S_t \underset{\sim}{W}_t$$

$$\underset{\sim}{X}_{t+1} = F_t \underset{\sim}{X}_t + \underset{\sim}{e}_t + T_t \underset{\sim}{V}_t$$

where

$\underset{\sim}{d}_t$ is $(w \times 1)$ deterministic

$\underset{\sim}{e}_t$ is $(r \times 1)$ deterministic

S_t is $(w \times w)$ deterministic

T_t is $(r \times v)$ deterministic

- allows correlation <u>within</u> <u>each</u> equation

eg

$$\underset{\sim}{Y}_t = G_t \underset{\sim}{X}_t + \underset{\sim}{d}_t + H_t \underset{\sim}{Z}_t$$

$$\underset{\sim}{X}_{t+1} = F_t \underset{\sim}{X}_t + \underset{\sim}{e}_t + K_t \underset{\sim}{Z}_t$$

where $\underset{\sim}{Z}_t$ is $(W+V) \times 1$ $WN(0, \{R_t\})$

H_t is $W \times (W+V)$ deterministic

K_t is $V \times (W+V)$ deterministic

- allows correlation <u>between</u> observation and state equations.

NOTE From the state equation

$$\underset{\sim}{X}_t = F_t \underset{\sim}{X}_{t-1} + \underset{\sim}{V}_{t-1}$$

$$= F_t \left(F_{t-1} \underset{\sim}{X}_{t-2} + \underset{\sim}{V}_{t-2} \right) + \underset{\sim}{V}_{t-1}$$

$$\vdots$$

$$= F_t F_{t-1} \ldots F_1 X_1 +$$
$$F_{t-1} \ldots F_2 V_1 +$$
$$F_{t-2} \ldots F_3 V_2 +$$
$$\vdots$$
$$F_{t-1} \underset{\sim}{V}_{t-2} + \underset{\sim}{V}_{t-1}$$

$$= f_t \left(\underset{\sim}{X}_1, \underset{\sim}{V}_1, \ldots \underset{\sim}{V}_{t-1} \right)$$

Similarly

$$\underset{\sim}{Y}_t = g_t \left(\underset{\sim}{X}_1, \underset{\sim}{V}_1, \ldots, \underset{\sim}{V}_{t-1}, \underset{\sim}{W}_t \right)$$

NOTE For $1 \le s \le t$

$$E[\underset{\sim}{V}_t \underset{\sim}{X}_s^T] = E[\underset{\sim}{V}_t \underset{\sim}{Y}_s^T] = 0$$

$$E[\underset{\sim}{W}_t \underset{\sim}{X}_s^T] = E[\underset{\sim}{W}_t \underset{\sim}{Y}_s^T] = 0$$

similar to the univariate case.

The state equation describes the evolution in time of the unobserved "driving" process $\{\underset{\sim}{X}_t\}$, which in turn causes the variation in observation $\{\underset{\sim}{Y}_t\}$

This is a useful model for many physical systems.

EXAMPLE $\{Y_t\} \sim AR(1)$

$$Y_t = \phi Y_{t-1} + Z_t$$

$$\{Z_t\} \sim WN(0, \sigma^2)$$

Let $X_{t+1} = \phi X_t + V_t \qquad (F_t \equiv \phi)$

$$X_1 = Y_1 = \sum_{j=0}^{\infty} \phi^j Z_{1-j}$$

$$V_t = Z_{t+1}$$

and $Y_t = X_t \qquad (G_t \equiv 1, \ \underset{\sim}{W}_t \equiv 0)$

EXAMPLE $\{Y_t\} \sim ARMA(1,1)$

Suppose

$$Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, where $|\phi|, |\theta| < 1$ so that $\{Y_t\}$ is causal and invertible.

i.e

$$Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$$

Let $\{X_t\}$ be defined by

$$\phi(B) X_t = Z_t$$

so that $\{X_t\}$ is causal AR(1).

Then $$\begin{bmatrix} X_t \\ X_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_t \end{bmatrix} + \begin{bmatrix} 0 \\ Z_{t+1} \end{bmatrix}$$

Letting

$$\underset{\sim}{X}_t = \begin{bmatrix} X_{t-1} \\ X_t \end{bmatrix}$$

$$\underset{\sim}{X}_1 = \begin{bmatrix} \sum_{j=0}^{\infty} \phi^j Z_{-j} \\ \sum_{j=0}^{\infty} \phi^j Z_{1-j} \end{bmatrix}$$

we have that

$$Y_t = G_t \underset{\sim}{X}_t \quad \text{where } G_t = [0, 1]$$

as $G_t \underset{\sim}{X}_t = \theta X_{t-1} + X_t$

$$= \theta X_{t-1} + \phi X_{t-1} + Z_t$$

Need to substitute in $X_t = \phi X_{t-1} + Z_t$.

$$Y_t = \theta X_{t-1} + X_t \qquad (1)$$

$$= \theta(\phi X_{t-2} + Z_{t-1}) + (\phi X_{t-1} + Z_t)$$

$$= \phi(X_{t-1} + \theta X_{t-2}) + \theta Z_{t-1} + Z_t$$

$$= \phi Y_{t-1} + \theta Z_{t-1} + Z_t \qquad \text{by (1)}$$

$\therefore \{Y_t\} \sim \text{ARMA}(1,1)$ has the state space representation

$$Y_t = G_t \underset{\sim}{X}_t \qquad G_t = [0, 1]$$

$$\underset{\sim}{X}_{t+1} = F_t \underset{\sim}{X}_t + \underset{\sim}{V}_t \qquad F_t = \begin{bmatrix} 0 & 1 \\ 0 & \phi \end{bmatrix}$$

where $Z_t \sim WN(0, \sigma^2)$

$$\underset{\sim}{X}_t = \begin{bmatrix} X_{t-1} \\ X_t \end{bmatrix} \qquad \underset{\sim}{V}_t = \begin{bmatrix} 0 \\ Z_{t+1} \end{bmatrix}$$

EXAMPLE MA(1)

Let $\{Y_t\} \sim MA(1)$

$$Y_t = Z_t + \theta Z_{t-1}$$

Where $\{Z_t\} \sim WN(0, \sigma^2)$

Define $\underset{\sim}{X}_t$ by

$$\underset{\sim}{X}_t = \begin{bmatrix} Y_t \\ \theta Z_t \end{bmatrix}$$

Then

$$Y_t = G_t \underset{\sim}{X}_t$$

with $G_t = [1, 0]$

$$\underset{\sim}{X}_{t+1} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \underset{\sim}{X}_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} Z_t$$

- extended version of state space form

Alternately, write

$$X_t = Z_{t-1}$$

Then

$$Y_t = 0 X_t + Z_t = G_t X_t + Z_t$$
$$X_{t+1} = Z_t = F_t X_t + Z_t$$

where $G_t = [0]$

$F_t = [0]$

Stationarity

Let

$$\underset{\sim}{Y}_t = G\underset{\sim}{X}_t + \underset{\sim}{W}_t \qquad t = 0, \pm 1, \pm 2, \ldots$$

$$\underset{\sim}{X}_{t+1} = F\underset{\sim}{X}_t + \underset{\sim}{V}_t \qquad t = 0, \pm 1, \pm 2, \ldots$$

i.e. $F_t \equiv F$ $v \times v$ constant matrix

$G_t \equiv G$ $w \times v$ constant matrix

$$\{\underset{\sim}{V}_t\} \sim WN(0, Q) \qquad v \times 1$$

$$\{\underset{\sim}{W}_t\} \sim WN(0, R) \qquad v \times 1$$

$$E[\underset{\sim}{V}_s \underset{\sim}{W}_t^T] = 0 \quad \forall s, t.$$

The state equation is **stable** if

$$|I - Fz| \neq 0$$

$\forall\ |z| \leq 1$

i.e all the eigenvalues of F lie **inside** the unit circle

Then

$$\underset{\sim}{X}_t = \sum_{j=0}^{\infty} F^j \underset{\sim}{V}_{t-j-1}$$

$$\underset{\sim}{Y}_t = \underset{\sim}{W}_t + \sum_{j=0}^{\infty} GF^j \underset{\sim}{V}_{t-j-1}$$

provide the **unique** stationary solution.

Note

The model is linear in $\underset{\sim}{X}_t$

"DYNAMIC LINEAR MODEL"

Because of this, estimation of the (unobserved) states is relatively straightforward.

We use the results from

- General Linear Model
- Multivariate Normal Distribution

to define a recursive calculation scheme

"KALMAN FILTER".

5.2 THE BASIC STRUCTURAL MODEL

Suppose

(I) $Y_t = M_t + W_t \qquad W_t \sim WN(0, \sigma_W^2)$

(II) $M_{t+1} = M_t + V_t \qquad V_t \sim WN(0, \sigma_V^2)$

with $M_1 = m_1$ a fixed constant

In the observation equation (I), the observed values Y_t follow a time-varying mean M_t process.

In the state equation (II), M_t follows a Random Walk.

$$D_t = Y_t - Y_{t-1}$$

$$= (M_t + W_t) - (M_{t-1} + W_{t-1})$$

$$= (M_t - M_{t-1}) + (W_t - W_{t-1})$$

$$= V_{t-1} + (W_t - W_{t-1})$$

Note that $E[D_t] = 0$

$$E[D_t D_{t+h}] = E[(V_{t-1} + (W_t - W_{t-1}))(V_{t+h-1} + (W_{t+h} - W_{t+h-1}))]$$

$$= E[V_{t-1} V_{t+h-1}] +$$
$$E[V_{t-1}(W_{t+h} - W_{t+h-1})] +$$
$$E[(W_t - W_{t-1}) V_{t+h-1}] +$$
$$E[(W_t - W_{t-1})(W_{t+h} - W_{t+h-1})]$$

$h = 0$

$$E[D_t^2] = E[V_{t-1}^2] + 0 + 0 + E[(W_t - W_{t-1})(W_t - W_{t-1})]$$

$$= \sigma_V^2 + 2\sigma_W^2$$

$h = 1$

$$= 0 + 0 + 0 + E[(W_t - W_{t-1})(W_{t+1} - W_t)]$$

$$= -\sigma_W^2$$

$h = 2, 3, \ldots$

$$E[D_t D_{t+h}] = 0 .$$

$$\therefore \quad \rho_D(h) = \begin{cases} \dfrac{-\sigma_w^2}{2\sigma_w^2 + \sigma_v^2} & |h| = 1 \\ 0 & \text{otherwise} \end{cases}$$

$\{D_t\} \sim MA(1)$

[redacted] is called the "local-level"

[redacted] we do not observe M_t
[redacted] may be the focus of interest

- [redacted] $y_1, \ldots, y_n$ we wish to
[redacted] $m_1, \ldots, m_n$ and σ_v^2, σ_w^2.

LOCAL LINEAR MODEL

(I) $Y_t = M_t + W_t$

(II) $M_t = M_{t-1} + \underbrace{B_{t-1}}_{\text{LOCAL TREND}} + V_{t-1}$

- could have B_{t-1} deterministic
- could have

$$B_t = B_{t-1} + U_{t-1}$$

- random walk model

Let $\underset{\sim}{X}_t = \begin{bmatrix} M_t \\ B_t \end{bmatrix}$ $\underset{\sim}{V}_t = \begin{bmatrix} V_t \\ U_t \end{bmatrix}$

Then

$$\underset{\sim}{X}_{t+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \underset{\sim}{X}_t + \underset{\sim}{V}_t$$

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \underset{\sim}{X}_t + W_t$$

i.e $\begin{matrix} V=2 \\ W=1 \end{matrix}$

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \sigma_v^2 & \bullet \\ 0 & \sigma_w^2 \end{bmatrix} \qquad R = \sigma_w^2$$

5.3 State Space Representation of ARIMA

Causal AR(p)

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p+1} = Z_{t+1}$$

Let

$$\underset{\sim}{X}_t = \begin{bmatrix} Y_{t-p+1} \\ \vdots \\ Y_t \end{bmatrix}$$

Then $Y_t = [0\ 0\ 0\ \cdots\ 1]\ \underset{\sim}{X}_t$

$$\underset{\sim}{X}_{t+1} = \Phi_p \underset{\sim}{X}_t + \underset{\sim}{V}_t$$

$$\Phi_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \phi_p & \phi_{p-1} & \cdots & & \phi_1 \end{bmatrix} \qquad \underset{\sim}{V}_t = \begin{bmatrix} 0 \\ \vdots \\ Z_{t+1} \end{bmatrix}$$

NOTE As $\{Y_t\}$ is causal this process is stable.

CAUSAL ARMA(p,q)

$$\phi(B) Y_t = \theta(B) Z_t \qquad Z_t \sim WN(0, \sigma^2)$$

Let

$$r = \max(p, q+1)$$

$$\phi_j = 0 \quad j > p$$

$$\theta_j = 0 \quad j > q$$

$$\theta_0 = \phi_0 = 1.$$

Define $\{U_t\}$ by

$$\phi(B) U_t = Z_t$$

- $\{U_t\}$ is causal AR, and

$$Y_t = \theta(B) U_t$$

Let $\underset{\sim}{X}_t = \begin{bmatrix} U_{t-r+1} \\ U_{t-r+2} \\ \\ \\ U_t \end{bmatrix}$

Then $\underset{\sim}{X}_{t+1} = \underset{\sim}{\Phi}_r \underset{\sim}{X}_t + \underset{\sim}{V}_t$ (all $r \times 1$)

$$Y_t = [\theta_{r-1} \cdots \theta_1 \ \theta_0] \underset{\sim}{X}_t$$

ARMA(p, d, q)

IF $Y_t \sim ARMA(p, d, q)$ then

$$\nabla^d Y_t = (1-B)^d Y_t \sim ARMA(p, q)$$

Thus

$$\nabla^d Y_t = G \underset{\sim}{X}_t$$

where

$$\underset{\sim}{X}_{t+1} = F \underset{\sim}{X}_t + \underset{\sim}{V}_t$$

where $F = \Phi_r$

$$G = [\theta_{r-1} \ldots \theta_1 \theta_0]$$

$$\underset{\sim}{V}_t = [0 \ldots 0 Z_{t+1}]^T$$

Now

$$\nabla^d Y_t = (1-B)^d Y_t$$

$$= \sum_{j=0}^{d} \binom{d}{j} (-1)^j Y_{t-j}$$

$$\therefore \quad Y_t = \nabla^d Y_t - \sum_{j=1}^{d} \binom{d}{j} (-1)^j Y_{t-j}$$

Let $\underset{\sim}{Y}_{t-1} = \begin{bmatrix} Y_{t-d} \\ \vdots \\ Y_{t-1} \end{bmatrix}$

Then

$$\underset{\sim}{Y}_t = AG\underset{\sim}{X}_t + B\underset{\sim}{Y}_{t-1}$$

where

$$A = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ (-1)^{d+1}\binom{d}{d} & & \cdots & & d \end{bmatrix}$$

A: $d \times 1$ $\quad$ B: $d \times d$

Final row of B is B_d,

$$\left[(-1)^{d+1}\binom{d}{d} \quad (-1)^{d}\binom{d}{d-1} \quad \cdots \quad (-1)^{2}\binom{d}{1}\right]$$

Let

$$\underset{\sim}{T}_t = \begin{bmatrix} \underset{\sim}{X}_t \\ \underset{\sim}{Y}_{t-1} \end{bmatrix}$$

Then we have the following state space model in $(Y_t, \underset{\sim}{T}_t)$

I) $$Y_t = \begin{bmatrix} G & B_d \end{bmatrix} \underset{\sim}{T}_t$$

II) $$\underset{\sim}{T}_{t+1} = \begin{bmatrix} F & 0 \\ AG & B \end{bmatrix} \underset{\sim}{T}_t + \begin{bmatrix} \underset{\sim}{V}_t \\ \underset{\sim}{0} \end{bmatrix}$$

with $$\underset{\sim}{T}_1 = \begin{bmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{Y}_0 \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{\infty} F^j \underset{\sim}{V}_{-j} \\ \underset{\sim}{Y}_0 \end{bmatrix}$$

5.4 THE KALMAN FILTER

The Kalman Filter is an efficient computation tool for solving inference problems for state space models:

Three classic problems:

PREDICTION: Computation of

$$E[\underset{\sim}{X}_t \mid \underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_{t-1}]$$

(one step ahead optimal forecasting

FILTERING Computation of

$$E[\underset{\sim}{X}_t \mid \underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_t]$$

Smoothing Computation of

$$E[\underset{\sim}{X}_t \mid \underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_n]$$

for $n > t$

Filtering - "prospective"

Smoothing - "retrospective"

Let $\mathcal{Y}_t = (\underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_t)^T$

- information set available at time t.

NOTE Let $\underset{\sim}{X} = (X_1, \ldots, X_v)^T$

and

$$P_t(\underset{\sim}{X}) = (P_t(X_1), \ldots, P_t(X_v))^T$$

where

$$P_t(X_i) = P(X_i \mid \mathcal{Y}_t)$$

is the best linear predictor of X_i in terms of $\mathcal{Y}_t = (\underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_t)^T$.

In the Gaussian case

$$P_t(\underset{\sim}{X}) = E[\underset{\sim}{X} \mid \underset{\sim}{Y}_1 \ldots \underset{\sim}{Y}_t]$$

$P_t(\cdot)$ is a linear operator

$$P_t(A\underset{\sim}{X}) = AP_t(\underset{\sim}{X})$$

$$P_t(\underset{\sim}{X}_1 + \underset{\sim}{X}_2) = P_t(\underset{\sim}{X}_1) + P_t(\underset{\sim}{X}_2)$$

$P_t(\cdot)$ is an orthogonalizing projection operator i.e.

$$(\underset{\sim}{X} - P_t(\underset{\sim}{X}))^T \underset{\sim}{Y}_s = 0$$

for $s = 0, 1, \ldots, t$.

By definition, $P_t(\cdot)$ is a linear combination of $\underset{\sim}{Y}_0, \ldots, \underset{\sim}{Y}_t$

Thus for the three key problems identified previously, we wish to compute

PREDICTION: $P_{t-1}(\underset{\sim}{X}_t)$

FILTERING: $P_t(\underset{\sim}{X}_t)$

SMOOTHING: $P_n(\underset{\sim}{X}_t) \qquad n > t.$

The Kalman Filter performs recursive calculations of these quantities.

PLAN OF ALGORITHM

"FILTER FORWARD, SMOOTH BACKWARD"

i.e

① Compute

$$P_{t-1}(\underset{\sim}{X}_t), \quad P_t(\underset{\sim}{X}_t)$$

moving from $t = 1, 2, \ldots n$

"FORWARDS"

② Compute

$$P_n(\underset{\sim}{X}_t)$$

for $t = n-1, n-2, \ldots 1.$

"BACKWARDS"

Distributional version

Let $y_0, \ldots, y_n$

$x_0, \ldots, x_n$

denote realized observations / states.

Full likelihood:

$$f(\underset{\sim}{x}, y) = \prod_{t=1}^{n} f_t(y_t \mid x_t)\, p_t(x_t \mid x_{t-1})$$

Prediction

$$p_{t+1}(x_{t+1} \mid y_0, \ldots, y_t)$$

Filtering

$$p_t(x_t \mid y_0, \ldots, y_t)$$

Now, by Bayes Theorem

$$p_t(x_t \mid y_0, \ldots, y_{t-1}, y_t)$$

$$= \frac{f_t(y_t \mid x_t)\, p_t(x_t \mid y_0, \ldots, y_{t-1})}{\int f_t(y_t \mid x_t)\, p_t(x_t \mid y_0, \ldots, y_{t-1})\, dx_t}$$

Denominator is

$$f_t(y_t \mid y_0, \ldots, y_{t-1})$$

$$p_{t+1}(x_{t+1} \mid y_0, \ldots, y_t)$$

$$= \int p_{t+1}(x_{t+1} \mid x_t)\, p_t(x_t \mid y_0, \ldots, y_t)\, dx_t.$$

Thus

Filtering is a posterior distribution calculation

Prediction is a posterior predictive distribution calculation.

SEQUENTIAL PROCESSING.

1. Obtain y_1

   - compute $p_1(x_1 / y_0, y_1)$

   - compute $p_2(x_2 / y_0, y_1)$

2. obtain y_2

   - compute $p_2(x_2 / y_0, y_1, y_2)$ using $p_2(x_2 / y_0, y_1)$

   - compute $p_3(x_3 / y_0, y_1, y_2)$ using $p_2(x_2 / y_0, y_1, y_2)$

etc.

In the Gaussian case, all calculations

posterior
posterior predictive
normalization constant

are available in <u>closed form</u>.

Calculations are essentially given by

"BAYES ESTIMATES FOR THE LINEAR MODEL"

by Lindley + Smith (1972)

Suppose

$$\underset{\sim}{Y}_t = G\underset{\sim}{X}_t + \underset{\sim}{W}_t$$

$$\underset{\sim}{X}_{t+1} = F\underset{\sim}{X}_t + \underset{\sim}{V}_t$$

$$\underset{\sim}{W}_t \sim N(0, \Sigma_W)$$

$$\underset{\sim}{V}_t \sim N(0, \Sigma_V)$$

} uncorrelated.

Then

$$\underset{\sim}{Y}_t \sim N(G\underset{\sim}{X}_t, \Sigma_W)$$

$$\underset{\sim}{X}_{t+1} \sim N(F\underset{\sim}{X}_t, \Sigma_V)$$

Suppose this holds for $t = 1, 2, \ldots$

$t=1$

a) $p(\underset{\sim}{X}_1 \mid \underset{\sim}{Y}_0, \underset{\sim}{Y}_1) \propto p(\underset{\sim}{Y}_1 \mid \underset{\sim}{X}_1)\, p(\underset{\sim}{X}_1 \mid \underset{\sim}{X}_0)$

b) $p(\underset{\sim}{X}_2 \mid \underset{\sim}{Y}_0, \underset{\sim}{Y}_1) = \int p(\underset{\sim}{X}_2 \mid \underset{\sim}{X}_1)\, p(\underset{\sim}{X}_1 \mid \underset{\sim}{Y}_0, \underset{\sim}{Y}_1)\, d\underset{\sim}{X}_1$

For (a)

$$p(\underset{\sim}{Y}_1 \mid \underset{\sim}{X}_1) \propto \exp\left\{-\tfrac{1}{2}(\underset{\sim}{Y}_1 - G\underset{\sim}{X}_1)^T \Sigma_W^{-1} (\underset{\sim}{Y} - G\underset{\sim}{X}_1)\right\}$$

$$p(\underset{\sim}{X}_1 \mid \underset{\sim}{X}_0) \propto \exp\left\{-\tfrac{1}{2}(\underset{\sim}{X}_1 - F\underset{\sim}{X}_0)^T \Sigma_V^{-1} (\underset{\sim}{X}_1 - F\underset{\sim}{X}_0)\right\}$$

To compute the posterior, we compute the sum of squares SSQ in the exponent by completing the square.

1.e

$$p(\underset{\sim}{X}_1 | \underset{\sim}{X}_0, \underset{\sim}{Y}_1) \propto \exp\left\{-\frac{1}{2} SSQ\right\}$$

$$SSQ = (\underset{\sim}{Y}_1 - G\underset{\sim}{X}_1)^T \Sigma_W^{-1} (\underset{\sim}{Y}_1 - G\underset{\sim}{X}_1) +$$

$$(\underset{\sim}{X}_1 - F\underset{\sim}{X}_0)^T \Sigma_V^{-1} (\underset{\sim}{X}_1 - F\underset{\sim}{X}_0)$$

$$= (\underset{\sim}{X}_1 - \underset{\sim}{m}_1)^T M_1 (\underset{\sim}{X}_1 - \underset{\sim}{m}_1) + c_1$$

[redacted].

Equating quadratic, linear and constant terms, we solve for

$$\underset{\sim}{m}_1, \; M_1, \; c_1$$

QUADRATIC TERM :

$$\underset{\sim}{x}_1^T M_1 \underset{\sim}{x}_1 = \underset{\sim}{x}_1^T G^T \Sigma_w^{-1} G \underset{\sim}{x}_1 + \underset{\sim}{x}_1^T \Sigma_v^{-1} \underset{\sim}{x}_1$$

$$\Rightarrow M_1 = G^T \Sigma_w^{-1} G + \Sigma_v^{-1}$$

LINEAR TERM :

$$\underset{\sim}{x}_1^T M_1 m_1 = \underset{\sim}{x}_1^T G^T \Sigma_w^{-1} \underset{\sim}{y}_1 + \underset{\sim}{x}_1^T \Sigma_v^{-1} F \underset{\sim}{x}_0$$

$$\therefore m_1 = M_1^{-1} \left(G^T \Sigma_w^{-1} \underset{\sim}{y}_1 + \Sigma_v^{-1} F \underset{\sim}{x}_0 \right)$$

$$= \left(G^T \Sigma_w^{-1} G + \Sigma_v^{-1} \right)^{-1} \left(G^T \Sigma_w^{-1} \underset{\sim}{y}_1 + \Sigma_v^{-1} F \underset{\sim}{x}_0 \right)$$

Constant term :

$$m_1^T M_1 m_1 + c_1 = \underset{\sim}{y}_1^T \Sigma_W^{-1} \underset{\sim}{y}_1 + \underset{\sim}{x}_0^T F^T \Sigma_V^{-1} F \underset{\sim}{x}_0$$

$$\therefore c_1 = \underset{\sim}{y}_1^T \Sigma_W^{-1} \underset{\sim}{y}_1 + \underset{\sim}{x}_0^T F^T \Sigma_V^{-1} F \underset{\sim}{x}_0 - \underset{\sim}{m}_1^T M_1 \underset{\sim}{m}_1$$

where $\underset{\sim}{m}_1$, M_1 are given previously.

$$\therefore p(\underset{\sim}{x}_1 | \underset{\sim}{x}_0, \underset{\sim}{y}_1) \propto \exp\left\{-\frac{1}{2}(\underset{\sim}{x}_1 - \underset{\sim}{m}_1)^T M_1 (\underset{\sim}{x}_1 - \underset{\sim}{m}_1)\right\}$$

i.e $\underset{\sim}{x}_1 | \underset{\sim}{x}_0, \underset{\sim}{y}_1 \sim N(\underset{\sim}{m}_1, M_1^{-1})$

b) $p(\underset{\sim}{X}_2 \mid \underset{\sim}{X}_0, \underset{\sim}{Y}_1)$

$$= \int p(\underset{\sim}{X}_2 \mid \underset{\sim}{X}_1)\, p(\underset{\sim}{X}_1 \mid \underset{\sim}{X}_0, \underset{\sim}{Y}_1)\, d\underset{\sim}{X}_1 .$$

$$p(\underset{\sim}{X}_2 \mid \underset{\sim}{X}_1) \propto \exp\left\{-\frac{1}{2}(\underset{\sim}{X}_2 - F\underset{\sim}{X}_1)^T \Sigma_V^{-1} (\underset{\sim}{X}_2 - F\underset{\sim}{X}_1)\right\}$$

$$p(\underset{\sim}{X}_1 \mid \underset{\sim}{X}_0, Y_1) \propto \exp\left\{-\frac{1}{2}(\underset{\sim}{X}_1 - \underset{\sim}{m}_1)^T M_1 (\underset{\sim}{X}_1 - \underset{\sim}{m}_1)\right\}.$$

∴ Integrand is again of the form

$$\exp\left\{-\frac{1}{2} SSQ\right\}$$

where $SSQ = (\underset{\sim}{X}_2 - F\underset{\sim}{X}_1)^T \Sigma_V^{-1} (\underset{\sim}{X}_2 - F\underset{\sim}{X}_1)$

$$+ (\underset{\sim}{X}_1 - \underset{\sim}{m}_1)^T M_1 (\underset{\sim}{X}_1 - \underset{\sim}{m}_1)$$

∴ Can again complete the square in $\underset{\sim}{x}_2$ to allow the computation to be performed

We have

$$SSQ = (\underset{\sim}{x}_1 - \underset{\sim}{m}_{12})^T M_{12} (\underset{\sim}{x}_1 - \underset{\sim}{m}_{12}) + c_{12}$$

where

$$M_{12} = F^T \Sigma_v^{-1} F + M_1$$

$$\underset{\sim}{m}_{12} = M_{12}^{-1} (F^T \Sigma_v^{-1} \underset{\sim}{x}_2 + M_1 \underset{\sim}{m}_1)$$

$$c_{12} = \underset{\sim}{x}_2^T \Sigma_v^{-1} \underset{\sim}{x}_2 + \underset{\sim}{m}_1^T M_1 m_1 - \underset{\sim}{m}_{12}^T M_{12} \underset{\sim}{m}_{12}$$

∴ can integrate out $\underset{\sim}{X}_1$ to obtain posterior predictive

$$p(\underset{\sim}{X}_2 | \underset{\sim}{X}_0, \underset{\sim}{Y}_1) \propto \exp\left\{-\frac{1}{2} c_{12}\right\}$$

$$= \exp\left\{-\frac{1}{2} (\underset{\sim}{X}_2 - \underset{\sim}{m}^*_{12})^T M^*_{12} (\underset{\sim}{X}_2 - \underset{\sim}{m}^*_{12})\right\}$$

where

$$M^*_{12} = \Sigma_v^{-1} - \Sigma_v^{-1} F \left(F^T \Sigma_v^{-1} F + M_1\right)^{-1} F^T \Sigma_v^{-1}$$

etc

i.e $\underset{\sim}{X}_2 | \underset{\sim}{X}_0, \underset{\sim}{Y}_1 \sim N(m^*_{12}, M^{*-1}_{12})$

In the Gaussian case

$$\underset{\sim}{Y}_t \sim N(G_t \underset{\sim}{X}_t, \Sigma_t)$$

$$\underset{\sim}{X}_{t+1} \sim N(F_t \underset{\sim}{X}_t, \Omega_t)$$

yields

$$p(\underset{\sim}{X}_t | \underset{\sim}{X}_0, \underset{\sim}{Y}_t) \equiv N(\underset{\sim}{a}_{t|t}, P_{t|t})$$

We use conditional expectation results

$$E[Y] = E_X[E_{Y|X}[Y|X=x]]$$

$$\text{Var}[Y] = E_X[\text{Var}_{Y|X}[Y|X=x]] + \text{Var}_X[E_{Y|X}[Y|X=x]]$$

to compute other results.

$$E[\underset{\sim}{X}_{t+1} | \underset{\sim}{X}_0, Y_t] = F_t \underset{\sim}{a}_{t|t}$$

$$\text{Var}[\underset{\sim}{X}_{t+1} | \underset{\sim}{X}_0, Y_t] = F_t P_{t|t} F_t^T + \Omega_t .$$

Due to properties of Gaussian distribution

$$p(\underset{\sim}{X}_{t+1} | \underset{\sim}{X}_0, Y_t) \equiv N(F_t \underset{\sim}{a}_{t|t}, F_t P_{t|t} F_t^T + \Omega_t)$$

$$\equiv N(\underset{\sim}{a}_{t+1|t}, P_{t+1|t})$$

- one step ahead predictive for $\underset{\sim}{X}_{t+1}$ given Y_t and $\underset{\sim}{X}_0$.

Again using conditional expectation

$$E[\underset{\sim}{Y}_{t+1} \mid X_0, \underset{\sim}{Y}_t] = G_{t+1}\underset{\sim}{a}_{t+1|t}$$

$$\text{Var}[\underset{\sim}{Y}_{t+1} \mid X_0, \underset{\sim}{Y}_t] = G_{t+1}P_{t+1|t}G_{t+1}^T + \Sigma_{t+1}$$

$$\therefore \underset{\sim}{Y}_{t+1} \mid X_0, Y_t \sim N(G_{t+1}\underset{\sim}{a}_{t+1|t}, G_{t+1}P_{t+1|t}G_{t+1}^T + \Sigma_{t}$$

- one step ahead predictive for

$\underset{\sim}{Y}_{t+1}$ given X_0, Y_t

$\therefore$ Starting with $a_{1|0}, P_{1|0}$ we have a complete recursion system.

If $\underset{\sim}{\theta}$ denotes the system parameters, then the likelihood $L(\underset{\sim}{\theta} | \underset{\sim}{y}_n)$ is given by

$$L(\underset{\sim}{\theta} | \underset{\sim}{y}_n) = \prod_{t=2}^{n} f_t(\underset{\sim}{y}_t | \underset{\sim}{y}_{t-1}, \underset{\sim}{\theta})$$

In the Gaussian case,

$$f_t(\underset{\sim}{y}_t | \underset{\sim}{y}_{t-1}, \underset{\sim}{\theta})$$

is also Gaussian for each t

$\therefore$ full likelihood is Gaussian, and evaluation is possible via the Kalman Filter recursions.

In the non-Gaussian case, we can form the same likelihood and perform approximate inference

"GAUSSIAN ESTIMATION"
"QUASI-LIKELIHOOD"

- again given by the Kalman Filter recursions

In the stable case with homoscedastic independent errors

$$\Sigma_t = \Sigma$$
$$\Omega_t = \Omega$$

estimation is consistent even in the non-Gaussian case.

Missing Data

Full Data

$$\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n$$

Often only a subset of the full data are actually observed.

$$\underset{\sim}{Y}_{i_1}, \ldots \underset{\sim}{Y}_{i_r}$$

with $\{i_j : j = 1, \ldots r\}$

an (irregularly spaced) index set.

To implement the Kalman Filter, define $\{\underset{\sim}{Y}_t^*\}$ by

$$\underset{\sim}{Y}_t^* = G_t^* \underset{\sim}{X}_t + \underset{\sim}{W}_t^* \qquad t=1,2,\ldots$$

where

$$G_t^* = \begin{cases} G_t & \text{if } t \in \{i_1, \ldots i_r\} \\ 0 & \text{otherwise} \end{cases}$$

$$\underset{\sim}{W}_t^* = \begin{cases} \underset{\sim}{W}_t & \text{if } t \in \{i_1, \ldots, i_r\} \\ \underset{\sim}{N}_t & \text{otherwise} \end{cases}$$

where $\underset{\sim}{N}_t \sim N(0, I_w)$

is orthogonal to $\underset{\sim}{X}_1$ and $\{\underset{\sim}{W}_t, \underset{\sim}{V}_t\}$.

i.e. for $t \notin \{i_1, \ldots i_r\}$

$\underset{\sim}{Y}_t^*$ is a value that is independent of all other $\underset{\sim}{Y}_t^*$ values

If $\underset{\sim}{y}_t^* = \begin{cases} \underset{\sim}{y}_t & \text{if } t \in \{i_1, \ldots i_r\} \\ 0 & \text{otherwise} \end{cases}$

represents the augmented observed series, then

$$L(\underset{\sim}{\theta}; \underset{\sim}{y}_{i_r}) \propto L(\underset{\sim}{\theta}, \underset{\sim}{y}_n^*)$$

$\underset{\sim}{y}_{i_r}$: data with missing obs

$\underset{\sim}{y}_n^*$: augmented data

Thus if we run the Kalman filter on the augmented series $\{\underset{\sim}{Y}_t^*\}$ we obtain recursions that estimate

$$\{\underset{\sim}{X}_t : t = 1, \ldots, n\}$$

and $\underset{\sim}{\theta}$ correctly.

- we use the dist^s of $\underset{\sim}{Y}_t \mid X_0, Y_{t-1}$ via $E[\underset{\sim}{Y}_t \mid X_0, Y_{t-1}]$ to fill in the missing values.

To fill in the missing values in a smoothing context

i.e to estimate $\underset{\sim}{Y}_t$ $t \notin \{i_1, \ldots, i_r\}$

from $\underset{\sim}{Y}_{i_1}, \ldots, \underset{\sim}{Y}_{i_r}$, we note that for such t,

$$E[\underset{\sim}{Y}_t \mid X_0, Y_r] = G_t E[\underset{\sim}{X}_t \mid X_0, Y_n^*]$$

$G_t \times$ smoothed estimate of $\underset{\sim}{X}_t$ given $\underset{\sim}{Y}_1^*, \ldots, \underset{\sim}{Y}_n^*$

i.e we again use the augmented Kalman Filter and its smoothed estimates.

THE E-M ALGORITHM

The Expectation - Maximization (EM) algorithm is a widely used approach for estimation of parameters in a likelihood setting when there are missing data.

In state-space models

$\underset{\sim}{Y}$ observed

$\underset{\sim}{X}$ "missing"/unobserved.

The EM algorithm is an iterative procedure for computing ML estimates for the "incomplete" data.

Let $\underset{\sim}{X}_n = \{\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n\}$

$\underset{\sim}{Y}_n = \{\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_n\}$

Let $\theta^{(0)}$ denote an initialization value for an iterative procedure that estimates the quantity $\hat{\theta}$ where

$$\hat{\theta} = \arg\max L(\theta; \underset{\sim}{Y}_n).$$

Let $\theta^{(i)}$ denote the estimate of θ obtained after i steps of the iterative procedure.

Let

$$L(\theta; Y_n) = \log L(\theta; Y_n)$$

$$L(\theta; X_n, Y_n) = \log L(\theta; X_n, Y_n)$$

denote the "incomplete" and "completed" log likelihoods respectively.

The procedure uses two steps

(i) E - step

(ii) M - step

(i) **E-step** Compute $Q(\theta|\theta^{(i)})$

$$Q(\theta|\theta^{(i)}) = E_{x|y}\left[\ell(\theta; x_n, y_n) \,\middle|\, y_n, \theta^{(i)}\right]$$

- The expected completed log-likelihood conditional on y_n and the current estimate $\theta^{(i)}$.

(ii) **M-step** Set

$$\theta^{(i+1)} = \arg\max_{\theta} Q(\theta|\theta^{(i)})$$

If the observed data are $\underset{\sim}{y}$ and the augmented data are $(\underset{\sim}{x}, \underset{\sim}{y})$, then

$$L(\theta; x_n, y_n) = \log f_{x,y}(\underset{\sim}{x}, \underset{\sim}{y} \mid \theta)$$

$$= \log f_{x|y}(\underset{\sim}{x} \mid \underset{\sim}{y}, \theta) + \log f_y(\underset{\sim}{y} \mid \theta)$$

and the conditional expectation in the E-step is with respect to

$$f_{x|y}(\underset{\sim}{x} \mid \underset{\sim}{y}, \theta^{(i)}) = \frac{f_{x,y}(\underset{\sim}{x}, \underset{\sim}{y} \mid \theta^{(i)})}{f_y(\underset{\sim}{y} \mid \theta^{(i)})}$$

Now, by definition

$$Q(\theta|\theta^{(i)}) = \int \log f_{X|Y}(\underset{\sim}{x}|\underset{\sim}{Y};\theta)\, f_{X|Y}(\underset{\sim}{x}|\underset{\sim}{Y},\theta^{(i)})\,\underset{\sim}{x}$$

$$+ L(\theta;\underset{\sim}{Y})$$

(second term is constant in the conditional expectation)

Differentiation with respect to θ yields

$$Q'(\theta|\theta^{(i)})$$

$$= \int \frac{\frac{\partial}{\partial\theta} f_{X|Y}(\underset{\sim}{x}|\underset{\sim}{Y};\theta)}{f_{X|Y}(\underset{\sim}{x}|\underset{\sim}{Y};\theta)}\, f_{X|Y}(\underset{\sim}{x}|\underset{\sim}{Y};\theta^{(i)})\, d\underset{\sim}{x}$$

$$+ L'(\theta;\underset{\sim}{Y})$$

Now, by definition

$$Q'(\theta^{(i+1)} | \theta^{(i)}) = 0$$

[$\theta^{(i+1)}$ maximizes Q]

$\therefore$ if $\{\theta^{(i)}\}$ converges to limit $\hat{\theta}$

then we must have

$$|\theta^{(i+1)} - \theta^{(i)}| < \varepsilon$$

for i large enough and so

$$|f_{X|Y}(\underset{\sim}{x} | \underset{\sim}{y}; \theta^{(i+1)}) - f_{X|Y}(\underset{\sim}{x} | \underset{\sim}{y}; \theta^{(i)})| < \varepsilon$$

$\therefore$ Solving for the Q' equation

$$0 = \int \frac{\partial}{\partial \theta} f_{X|Y}(\underset{\sim}{x} | \underset{\sim}{Y}; \theta)\Big|_{\theta=\hat{\theta}} \, d\underset{\sim}{x}$$

$$+ L'(\hat{\theta}; \underset{\sim}{Y})$$

But $\int \frac{\partial}{\partial \theta} f_{X|Y}(\underset{\sim}{x} | \underset{\sim}{Y}; \theta)\Big|_{\theta=\hat{\theta}} \, d\underset{\sim}{x} = 0$

by exchanging the order of integration and differentiation.

$\therefore L'(\hat{\theta}; \underset{\sim}{Y}) = 0$ and $\hat{\theta}$ is the ML estimate.

For state-space models we adopt a Gaussian likelihood approximation

We use the modified Q function

$$\tilde{Q}(\theta|\theta^{(i)}) = L(\theta; \hat{x}_n, y_n)$$

where $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$

where $\hat{x}_i = E[x_i | y_n, \theta^{(i)}]$

returned by the smoothing algorithm component of the Kalman Filter, with $\theta = \theta^{(i)}$.

For the M-step, standard ML can be used. (easy for ARMA)

6 FINANCIAL TIME SERIES MODELS

There is much current interest in models for financial time series

- stock prices
- foreign exchange indices
- composites / indices
- bond prices
- interest rates

Such series often exhibit characteristic patterns of behaviour.

Example Volatility

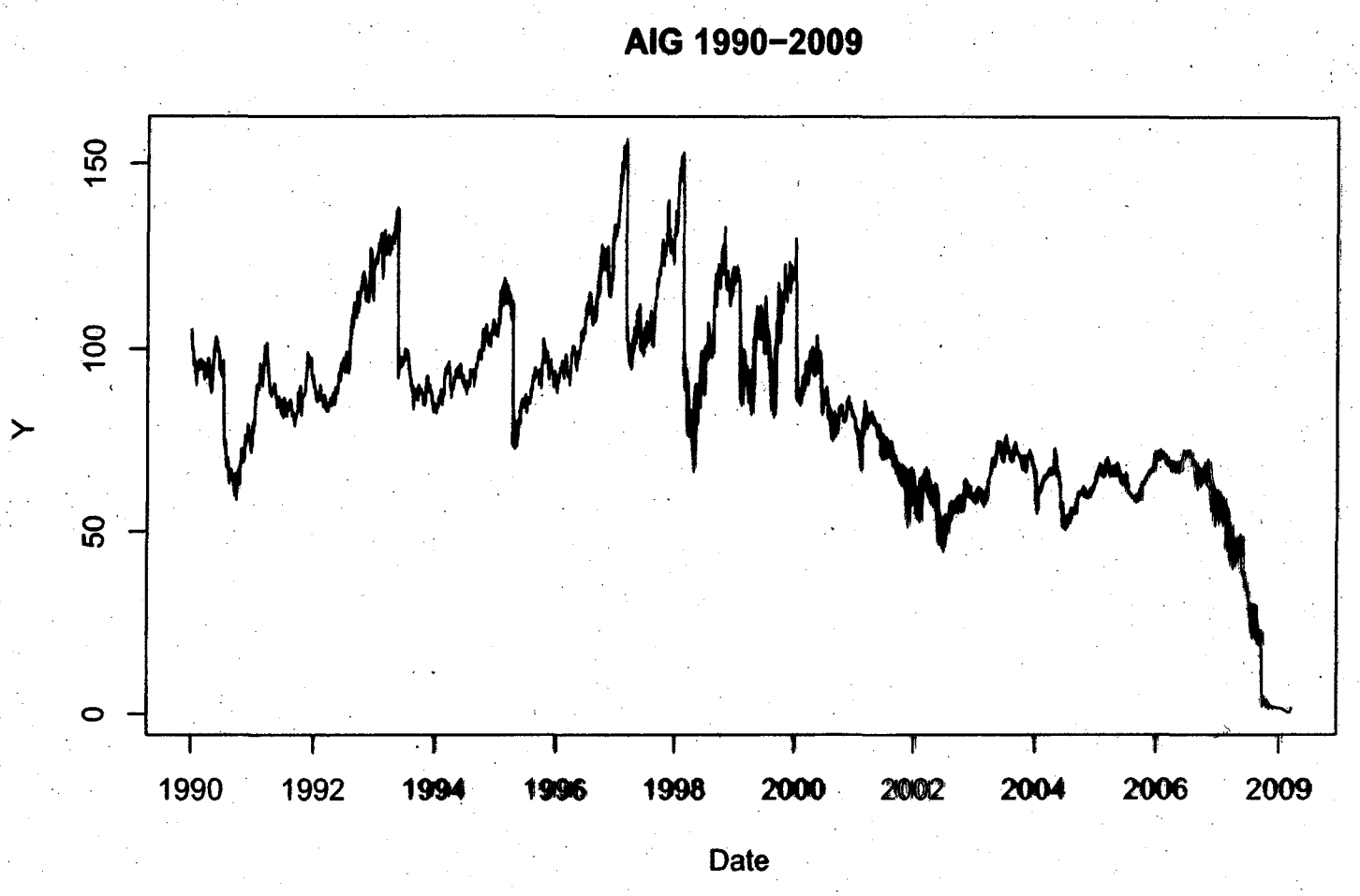
Volatility modelling is the modelling of a time-varying variance of an observed data series. Typically the series will exhibit

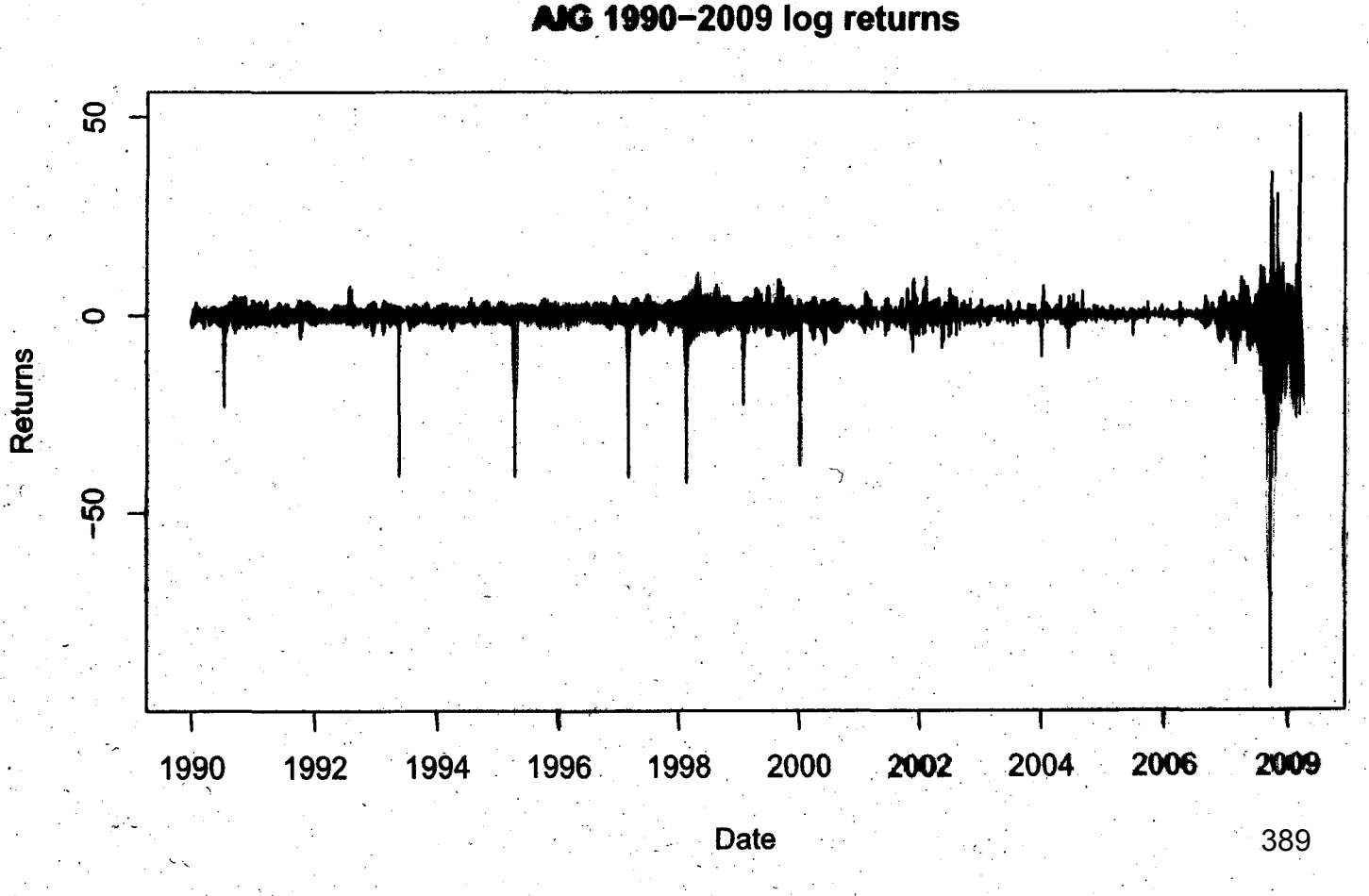
Conditional Heteroscedasticity

(time-varying conditional variance) and

Volatility Clustering

(conditional variances exhibit autocorrelation - periods of high and low volatility)


AIG 1990−2009
Y
150
100
50
0
1990
1992
1994
1996
1998
2000
2002
2004
2006
2009
Date


AIG 1990−2009 log returns
Returns
50
0
−50
1990
1992
1994
1996
1998
2000
2002
2004
2006
2009
Date

State space models are often used to model such processes.

A distinction is made between

PARAMETER-DRIVEN

and

OBSERVATION-DRIVEN

models.

The distinction is purely determined by the assumed conditional probability structure.

Let $\underset{\sim}{X}^{(t)} = (X_1, \ldots, X_t)^T$ STATES

$\underset{\sim}{Y}^{(t)} = (Y_1, \ldots, Y_t)^T$ OBS.

In a parameter-driven model

$$p(y_t \mid x_t, \underset{\sim}{x}^{(t-1)}, \underset{\sim}{y}^{(t-1)}) \equiv p(y_t \mid x_t)$$

$$p(x_{t+1} \mid x_t, \underset{\sim}{x}^{(t-1)}, \underset{\sim}{y}^{(t)}) \equiv p(x_{t+1} \mid x_t)$$

i.e. given x_t, (y_t, x_{t+1}) are conditionally independent of $(\underset{\sim}{x}^{(t-1)}, \underset{\sim}{y}^{(t-1)})$

so given $X_1 = x_1$

$$p(x_2, x_3, \ldots, x_n, y_1, \ldots, y_n \mid x_1)$$

$$= \left\{ \prod_{j=1}^{n} p(y_j \mid x_j) \right\} \left\{ \prod_{j=2}^{n} p(x_j \mid x_{j-1}) \right\}$$

In observation-driven models, again

$$p(y_t \mid x_t, \underset{\sim}{x}^{(t-1)}, \underset{\sim}{y}^{(t-1)}) = p(y_t \mid x_t)$$

but

$p(x_{t+1} \mid \underset{\sim}{y}^{(t)})$ is some specified conditional density for x_{t+1} formed as a function of $(y_1, \ldots, y_t)^T$, conditionally independent of $\underset{\sim}{x}^{(t)}$.

i.e.

$$p(x_1, \ldots, x_n, y_1, \ldots, y_n)$$

$$= \prod_{j=1}^{n} p(y_j \mid x_j) p(x_j \mid \underset{\sim}{y}^{(j-1)})$$

EXAMPLE PARAMETER-DRIVEN MODEL FOR COUNT DATA

Suppose for $t = 1, 2, \ldots$

(i) $Y_t \mid X_t = x_t \sim \text{Poisson}(\pi x_t)$

(ii) $$X_{t+1} \mid X_t = x_t = \begin{cases} 1 & x_{t+1} = a x_t \\ 0 & \text{otherwise} \end{cases}$$

i.e. $X_{t+1} = a X_t$

(iii) $X_1 \sim \text{Gamma}(\alpha, \lambda)$

The Gamma distribution is <u>conjugate</u> to the Poisson in the Bayesian calculation of

- posterior $p(x_t \mid y_t)$
- predictive $p(x_{t+1} \mid y_t)$
- predictive $p(y_{t+1} \mid y_t)$

We find that

$$p(x_t | y_t) \equiv \text{Gamma}(\alpha_t, \lambda_t)$$

$$p(x_{t+1} | y_t) \equiv \text{Gamma}(\alpha_t, \lambda_t / a)$$

$$p(y_{t+1} | y_t) \equiv \text{NegBin}\left(\alpha_t, 1 - \pi / \lambda_{t+1}\right)$$

where

$$\alpha_t = \alpha + \sum_{j=1}^{t} y_j$$

$$\lambda_t = \lambda a^{1-t} + \pi\left(\frac{1 - a^{-t}}{1 - a^{-1}}\right)$$

Here the Gamma distribution is in the standard parameterization

eg $$E[X_1] = \frac{\alpha}{\lambda}$$

The Negative Binomial is in the parameterization

$$Y \sim \text{NegBin}(\alpha, \lambda) \qquad \alpha > 0,\ 0 < \lambda < 1.$$

$$f_Y(y) = \binom{\alpha+y-1}{\alpha} (1-\lambda)^y \lambda^\alpha \qquad y = 0, 1, 2, \ldots$$

Also

$$E[Y_{t+1} | Y_t] = \frac{\alpha_t \pi}{\lambda_{t+1} - \pi}$$

$$\text{Var}[Y_{t+1} | Y_t] = \frac{\alpha_t \pi \lambda_{t+1}}{(\lambda_{t+1} - \pi)^2}$$

EXAMPLE AN OBSERVATION-DRIVEN MODEL

Suppose

$$p(x_{t+1} \mid y_t) = \begin{cases} 1 & x_{t+1} = y_t \\ 0 & \text{otherwise} \end{cases}$$

$$p(y_t \mid x_t) = f_Z(y_t - \phi x_t)$$

for some pdf f_Z

Then clearly

$$x_t - \phi x_t \sim Z_t \quad \text{for some } \{Z_t\}$$

$$\Rightarrow Y_t = \phi x_t + Z_t$$

$\therefore$ Defining $X_t = Y_{t-1} \Rightarrow X_t \sim AR(1)$.

OBSERVATION-DRIVEN MODELS BASED ON THE EXPONENTIAL FAMILY

Suppose

$$p(y_t|x_t) = \exp\{y_t x_t - b(x_t) + c(y_t)\}$$

i.e a natural exponential family distribution with canonical parameter x_t.

By standard results

$$\exp\{b(x_t)\} = \int \exp\{y_t x_t + c(y_t)\}\, dP_{y_t}$$

$$\dot{b}(x_t) = E[Y_t|X_t = x_t]$$

$$\ddot{b}(x_t) = \mathrm{Var}[Y_t|X_t = x_t]$$

Suppose

$$p(x_t \mid y_{t-1}) \equiv f(x_t \mid \alpha_{t|t-1}, \lambda_{t|t-1})$$

where

$$f(x \mid \alpha, \lambda) = \exp\{\alpha x - \lambda b(x) + A(\alpha, \lambda)\}$$

Then

$$p(x_t \mid y_{t-1}, y_t)$$

$$\propto p(y_t \mid x_t) f(x_t \mid y_{t-1})$$

$$= \exp\{y_t x_t - b(x_t) + c(y_t)\}$$

$$\times \exp\{\alpha_{t|t-1} x_t - \lambda_{t|t-1} b(x_t)\}$$

$$\propto \exp\{\alpha_{t|t} x_t - \lambda_{t|t} b(x_t)\}$$

$$\alpha_{t|t} = y_t + \alpha_{t|t-1} \qquad \lambda_{t|t} = 1 + \lambda_{t|t-1}.$$

$$\therefore\ p(x_t \mid y_t) \equiv f(x_t \mid \alpha_{t|t}, \lambda_{t|t})$$

Thus "prior" $f(x_t \mid \alpha_{t|t-1}, \lambda_{t|t-1})$

and "posterior" $f(x_t \mid \alpha_{t|t}, \lambda_{t|t})$

have the same functional form.

We merely update the parameters.

Using results for the exponential family it follows that

$$E[Y_{t+1} \mid Y_t] = \frac{\alpha_{t+1|t}}{\lambda_{t+1|t}}$$

Now suppose

$$\lambda_{t+1|t} = \delta \lambda_{t|t} = \delta(1 + \lambda_{t|t-1})$$

$$\alpha_{t+1|t} = \delta \alpha_{t|t} = \delta(y_t + \alpha_{t|t-1})$$

For some $0 < \delta < 1$. With these definitions, by recursion, it follows that

$$\lambda_{t+1|t} = \delta + \delta^2 + \ldots + \delta^{t-1} + \delta^t(1 + \lambda_{1|0})$$

$$\alpha_{t+1|t} = \delta y_t + \delta^2 y_{t-1} + \cdots + \delta^{t-1} y_2 + \delta^t(y_1 + \alpha_{1|0})$$

By the previous result

$$\hat{Y}_{t+1} = E[Y_{t+1} | Y_t] = \frac{\sum_{j=0}^{t-1} \delta^j y_{t-j} + \delta^{t-1} \alpha_{1|0}}{\sum_{j=0}^{t-1} \delta^j + \delta^{t-1} \lambda_{1|0}}.$$

In the limit as $t \to \infty$

$$\lambda_{t+1|t} = \frac{\delta}{1-\delta}$$

and for large t

$$\text{level}_t = \delta \sum_{j=0}^{t-1} \delta^j y_{t-j}$$

If we set $\lambda_{1|0} = \frac{\delta}{1-\delta}$

then

$$\hat{Y}_{t+1} = (1-\delta) y_t + \delta \hat{Y}_t$$

EXPONENTIAL SMOOTHING

with

$$\hat{Y}_1 = \frac{(1-\delta)}{\delta} \alpha_{1|0}.$$

NOTE Under these assumptions

$$\underbrace{f(x_t \mid \alpha_{t|t-1}, \lambda_{t|t-1})}_{\text{"PRIOR"}} \propto \underbrace{f^{\delta}(x_t \mid \alpha_{t-1|t-1}, \lambda_{t-1|t-1})}_{\text{"POSTERIOR"}}$$

$p(x_t \mid Y_t)$ $\qquad$ $p(x_t \mid Y_{t-1})$

i.e to develop the new "prior" we raise the current posterior to power δ

for $0 < \delta < 1$.

- this "flattens" the current posterior.

EXAMPLE MODELLING OVERDISPERSION USING OBSERVATION-DRIVEN MODELS

Suppose

(i) $p(x_t \mid y_{t-1}) \equiv f(x_t \mid \alpha_{t|t-1}, \lambda_{t|t-1})$

where $\alpha_{t|t-1}$ and $\lambda_{t|t-1}$ parameterize the first two moments, say, and depend on $p(x_1)$ and $y_1, \ldots, y_{t-1}$ as before.

(ii) $p(y_t \mid x_t) \sim \text{Poisson}(\exp\{x_t\})$

i.e $E[Y_t \mid X_t] = \exp\{X_t\}$

The system depends on the parameters δ, $\lambda_{1|0}$ and $\kappa_{1|0}$

$$L(\delta, \lambda_{1:0}, \kappa_{1:0}) = \prod_{t=1}^{n-1} p(y_{t+1} | Y_t)$$

where $p(y_{t+1} | Y_t) \equiv \text{NegBin}\left(\kappa_{t+1|t}, \frac{1}{1+\lambda_{t+1|t}}\right)$

If $f(x|\alpha, \lambda)$ follows the Gamma form

(usual conjugate Bayesian calculation)

Using recursion we can compute the likelihood easily.

Note that the stochastic X_t model induces an <u>increase</u> in variance of

Y_t, so we can obtain an over dispersed model. However $\{Y_t\}$ not i.i.d.

VOLATILITY MODELLING WITH PARAMETER AND OBSERVATION-DRIVEN MODELS

A standard model for a time-evolving financial process, $\{S_t\}$, utilizes the stochastic differential equation (SDE)

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t$$

where $\{W_t\}$ is a continuous time stochastic process with independent Gaussian increments (BROWNIAN MOTION)

If $Y_t = \log S_t$, then it follows that

$$dY_t = \mu dt + \sigma dW_t$$

A way to interpret this equation is via the discrete ("Euler") approximation

Where we replace

dY_t by $Y_{t+1} - Y_t$

dt by Δ

and imagine discretely observed values $Y_1, Y_2, \ldots Y_n$ at times $\Delta, 2\Delta, 3\Delta, \ldots$

1.e

$$Y_{t+1} - Y_t = \mu\Delta + \sigma Z_t$$

where $\{Z_t\} \sim NID(0,1)$

1.e $Y_{t+1} | Y_t \sim N(Y_t + \Delta\mu, \sigma^2)$

μ is the drift
σ is the volatility.

However this model is <u>not</u> appropriate for most financial series

- when we look at the observed series $\{y_{t+1} - y_t, t = 1, 2, \ldots\}$ it does not appear to be Gaussian.

The four stylized facts:

The series defined for $t = 2, 3, \ldots$ by

$$r_t = y_{t+1} - y_t = \log\left(\frac{s_{t+1}}{s_t}\right)$$

is termed the log returns series.

It is well known that for financial series the log returns have

(i) A leptokurtic (heavy-tailed) distribution

(ii) Time-varying variance that exhibits autocorrelation (volatility clustering)

(iii) Leverage: an asymmetric influence of increases in s_t on the variance

(iv) Covariation across different serie

A simple way to extend the initial model is to make σ time dependent

i.e $$dY_t = \mu\, dt + \sigma(t)\, dW_t$$

In the literature this has been achieved in two ways

① OBSERVATION-DRIVEN MODEL

ARCH / GARCH

AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC

G. GENERALIZED ARCH.

ARCH

Suppose $\mu = 0$, and let

$$Y_t \mid \mathcal{Y}_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2$$

GENERALIZED ARCH

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

- need $\sigma_t^2 > 0 \quad \forall t$; this puts restrictions on the parameters.

② PARAMETER-DRIVEN MODEL

STOCHASTIC VOLATILITY

Let

$$dY_t = \mu\,dt + \sqrt{V_t}\,dW_{1t}$$

$$dV_t = a(V_t,t)\,dt + b(V_t,t)\,dW_{2t}$$

where $a(V_t,t)$, $b(V_t,t)$ are deterministic functions of (V_t,t)

W_{1t}, W_{2t} - two correlated Gaussian white noise processes

i.e in a discrete approximation

$$Y_t \mid V_t = v_t \sim N(Y_{t-1} + \mu\Delta, v_t)$$

and $\{V_t\}$ follows a general "diffusion" process.

i.e $\{V_t\}$ is <u>stochastic</u>; its time evolution determined by the diffusion functions

$a(V_t, t)$ - drift

$b(V_t, t)$ - volatility.

A simpler discrete time model has

$$Y_t \mid h_t \sim N(0, \exp\{h_t\})$$

$$h_t \sim AR(1)$$

i.e $$Y_t = \exp\{h_t/2\} Z_{1t}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_h Z_{2t}$$

where $\{Z_{1t}\} \sim NID(0,1)$

$\{Z_{2t}\} \sim NID(0,1)$

where Z_{1t}, Z_{2t} are uncorrelated.

The stochastic volatility model is a

NON-LINEAR / NON GAUSSIAN

STATE SPACE MODEL

- The observed process $\{Y_t\}$ is modelled via the unobserved process $\{h_t\}$ or $\{V_t\}$.

- Kalman Filtering ideas can be used to make inference about the unobserved "states" (volatilities) and the system parameters.

7. SPECTRAL ANALYSIS

The spectral or frequency domain representation of time series $\{X_t\}$ gives an alternative approach to construction of and inference for time series models.

The representation focusses on stationary processes only. It can be shown that every stationary process can be represented by spectral function that is unique

This representation is a fundamental construction of stationary processes.

In signal processing, the frequency domain representation is commonly used -

- statisticians usually regard the "time domain" as paramount, as the data are normally collected in a time sequential fashion.

- a simple data transformation converts time domain data to the frequency domain for inspection and analysis.

7.1 Spectral Representations of Stationary Processes

A stationary process is characterized via its ACVF: if $\{X_t\}$ is zero mean

$$\gamma_X(h) = E[X_t X_{t+h}]$$

In fact, as γ_X is a non-negative definite function, it has a unique representation in terms of "frequency contributions"

DEFINITION THE SPECTRAL DENSITY FUNCTION

Let $\{X_t\}$ be a zero-mean stationary time series process with autocovariance sequence γ_X, and suppose

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$$

Then the spectral density function (SDF) of $\{X_t\}$ is the function $f_X(\cdot)$ defined by

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\lambda h} \gamma_X(h)$$

for $\lambda \in \mathbb{R}$.

NOTE

$$e^{ih\lambda} = \cos\lambda h + i\sin\lambda h$$

$$\therefore |e^{ih\lambda}| = \sqrt{\cos^2\lambda h + \sin^2\lambda h} = 1$$

and thus

$$f_X(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma_X(h)$$

is well-defined as

$$\sum_{h=-\infty}^{\infty} |e^{-ih\lambda}\gamma_X(h)| \leq \sum_{h=-\infty}^{\infty} |e^{-ih\lambda}|\,|\gamma_X(h)|$$

$$= \sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$$

by assumption.

NOTE Cos and sin are periodic functions

$\therefore$ $f_x(\lambda)$ is periodic

$\Rightarrow$ we can restrict attention to $(-\pi, \pi]$

PROPERTIES

(1) $f_x(\lambda) = f_x(-\lambda)$

$$f_x(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} [\cos \lambda h - i \sin \lambda h] \gamma_x(h)$$

$$= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos \lambda h \, \gamma_x(h) \quad \left[\begin{array}{c}\text{sin terms} \\ \text{cancel}\end{array}\right]$$

$$= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(-\lambda h) \gamma_x(h)$$

$$= f_x(-\lambda)$$

(iii) $f_X(\lambda)$ is real-valued if $\{X_t\}$ is real-valued.

(iv) $f_X(\lambda) \geq 0 \quad \forall \lambda.$

(iv) $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda)\, d\lambda$

$$= \int_{-\pi}^{\pi} \cos(h\lambda) f_X(\lambda)\, d\lambda$$

as $\int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda) d\lambda = \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{i(h-k)\lambda} \gamma_X(k) d\lambda$

$$= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) \int_{-\pi}^{\pi} e^{i(h-k)\lambda} d\lambda$$

$$= \gamma_X(h)$$

(only non zero term in sum is when $k=h$.

NOTE The summability condition

$$\sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty$$

is not a necessary condition for the existence of a SDF. If

(i) $f_x(\lambda) \geq 0 \quad \forall \lambda \in (0, \pi]$

(ii) $\gamma_x(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_x(\lambda)\, d\lambda \quad \forall h$

then f_x is the SDF for $\{X_t\}$

NOTE There is a 1-1 correspondence between ACVFs and SDFs.

i.e the SDF is unique to γ_x

THEOREM A real-valued function f defined on $(-\pi, \pi]$ is the SDF of a stationary process if and only if

(i) $f(\lambda) = f(-\lambda)$

(ii) $f(\lambda) \geq 0$

(iii) $\int_{-\pi}^{\pi} f(\lambda)\, d\lambda < \infty$

Proof

Suppose f satisfies (i), (ii), (iii).

Then the function

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda \text{ is } \underline{\text{even}}$$

Let $a_r \in \mathbb{R} \quad r = 1, \ldots, n$

Then

$$\sum_{r=1}^{n} \sum_{s=1}^{n} a_r \gamma(r-s) a_s$$

$$= \int_{-\pi}^{\pi} \sum_{r=1}^{n} \sum_{s=1}^{n} a_r a_s e^{i\lambda(r-s)} f(\lambda)\, d\lambda$$

$$= \int_{-\pi}^{\pi} \left| \sum_{r=1}^{n} a_r e^{i\lambda r} \right|^2 f(\lambda)\, d\lambda$$

$$\geq 0$$

$\therefore \gamma$ is non-negative definite

$\therefore \gamma$ is an acvf for a stationary process.

Conversely Suppose γ is the acvf for a stationary process

- if γ is absolutely summable, proof that f has the given properties given in part above.

- Full proof omitted

This result further characterizes acvfs
- γ_X must have a representation via a suitable f.

f_x is the Fourier Transform of γ_x

γ_x is the Inverse Fourier Transform of f_x

NOTE The properties of f_x are similar to those of a pdf; the only exception is that

$$\int_{-\pi}^{\pi} f_x(\lambda)\, d\lambda < \infty$$

not

$$\int_{-\pi}^{\pi} f_x(\lambda)\, d\lambda = 1 .$$

(plus the additional evenness).

Spectral Representation Theorem

Theorem A function γ

$$\gamma : \mathbb{Z} \longrightarrow \mathbb{R}$$

is an acvf if and only if there exists a right continuous, non-decreasing bounded function F on $[-\pi, \pi]$ with $F(-\pi) = 0$, and, $\forall h \in \mathbb{Z}$

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\lambda} \, dF(\lambda)$$

If the corresponding series is real-valued then F is symmetric

$$\int_{(a,b]} dF(x) = \int_{[-b,-a)} dF(x) .$$

F is a "generalized distribution function" on $[-\pi, \pi]$ as

$$\frac{F(\lambda)}{F(\pi)}$$

is a distribution function on $[-\pi, \pi]$

F is termed the <u>spectral distribution function</u>

Like ordinary distribution functions F can be purely discrete, or absolutely continuous, or mixed.

If F is purely discrete, then the corresponding density function f does not exist. If F is absolutely continuous

$$F(\lambda) = \int_{-\pi}^{\lambda} f(x)\,dx \qquad \forall \lambda \in (-\pi, \pi]$$

then f is the sdf.

EXAMPLE Suppose $X_t = A\cos\omega t + B\sin\omega t$

where A, B are uncorrelated

$$E[A] = E[B] = 0$$

$$\mathrm{Var}[A] = \mathrm{Var}[B] = 1.$$

Then it can be shown that

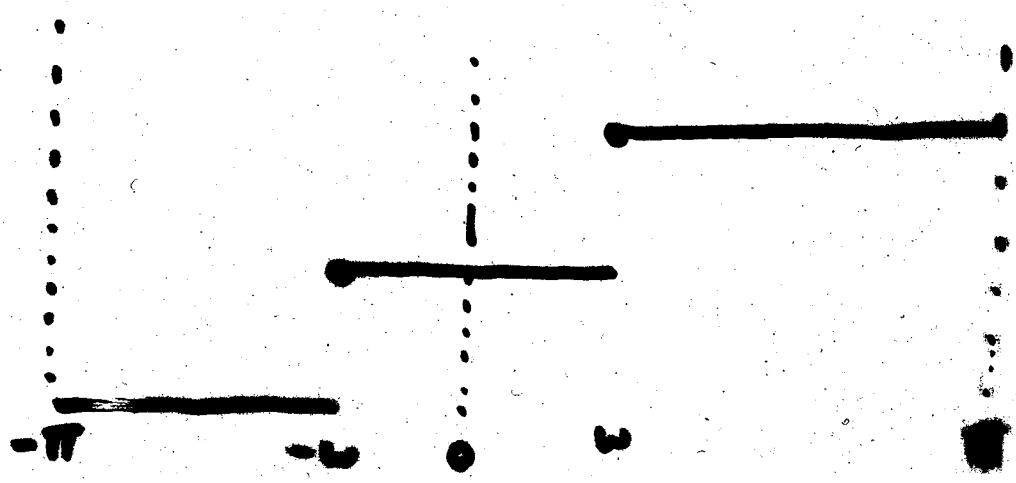
$$\gamma_X(h) = \cos \omega h \qquad h \in \mathbb{Z}.$$

Now

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = \sum_{h=-\infty}^{\infty} |\cos \omega h| = \infty$$

However if

$$F_X(\lambda) = \begin{cases} 0 & \lambda < -\omega \\ \frac{1}{2} & -\omega \leq \lambda \leq \lambda \\ 1 & \lambda \geq \omega \end{cases}$$



Then

$$\int_{(-\pi,\pi]} e^{ih\lambda}\, dF(\lambda)$$

$$= e^{-ih\omega} \cdot \frac{1}{2} + e^{ih\omega} \cdot \frac{1}{2}$$

$$= \frac{1}{2}\left[e^{ih\omega} + e^{-ih\omega} \right]$$

$$= \cos(h\omega)$$

$dF(\lambda)$ represents the instantaneous change in $F(\lambda)$

EXTENSION Suppose

$$X_t = \sum_{j=1}^{k} \left[A_j \cos(\omega_j t) + B_j \sin(\omega_j t) \right]$$

for $0 < \omega_1 < \omega_2 < \ldots < \omega_k < \pi$

where $A_1, \ldots A_k$ $B_1, \ldots, B_k$ are uncorrelated

$$E[A_j] = E[B_j] = 0$$

$$Var[A_j] = Var[B_j] = \sigma_j^2$$

Then

$$\gamma_X(h) = \sum_{j=1}^{k} \sigma_j^2 \cos(\omega_j h)$$

Then, by the above proof

$$F_X(\lambda) = \sum_{j=1}^{k} \sigma_j^2 F_j(\lambda)$$

where

$$F_j(\lambda) = \begin{cases} 0 & \lambda < -\omega_j \\ \frac{1}{2} & -\omega_j \leq \lambda \leq \omega_j \\ 1 & \lambda \geq \omega_j \end{cases}$$

Note

$$\text{Var}[X_t] = \gamma_X(0)$$

$$= \sum_{j=1}^{k} \sigma_j^2$$

This result shows that the variance of $\{X_t\}$ can be decomposed into the variances contributed by the sinusoids at the different frequencies

$$\omega_1, \omega_2, \ldots, \omega_k$$

It can be shown that <u>all</u> zero mean stationary processes have this kind of representation if we let

$$k \longrightarrow \infty$$

In the limit

$$X_t = \int_{(-\pi,\pi]} e^{it\lambda}\, dZ(\lambda) \qquad (*)$$

where

$$\{Z(\lambda) : -\pi < \lambda \leq \pi\}$$

is a stochastic process, so that $(*)$

is a stochastic integral

In general Z is complex-valued with uncorrelated increments

For finite k

$$dZ_k(\lambda) = \begin{cases} \frac{A_j + iB_j}{2} & \lambda = -\omega_j \quad j = 1, \ldots k \\ \frac{A_j - iB_j}{2} & \lambda = \omega_j \quad j = 1, \ldots k \\ 0 & \text{otherwise} \end{cases}$$

Here

$$E\left[dZ_k(\lambda)\, \overline{dZ_k(\lambda)}\right]$$

$$= \begin{cases} E\left[\frac{A_j^2 + B_j^2}{2}\right] = \frac{\sigma_j^2}{2} & \lambda = \pm\omega_j \\ 0 & \text{otherwise.} \end{cases}$$

NOTE

When we write

$$X_t = \int_{(-\pi,\pi]} e^{it\lambda}\, dZ(\lambda)$$

equality is in the mean-square sense

- both sides are random variables

  ($Z(\lambda)$ is a stochastic process on $(-\pi,\pi]$)

- $E\left[\left(X_t - \int_{(-\pi,\pi]} e^{it\lambda}\, dZ_k(\lambda)\right)^2\right] \to 0$ as $k \to \infty$

EXAMPLE $X_t \sim WN(0, \sigma^2)$

i.e $\gamma_X(0) = \sigma^2$

$\gamma_X(h) = 0 \qquad \forall\, h \neq 0$

Then

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma_X(h)$$

$$= \frac{1}{2\pi} \left[\cos(0\lambda) - i \sin(0\lambda) \right] \sigma^2$$

$$= \frac{\sigma^2}{2\pi} \qquad -\pi < \lambda \leq \pi$$

$$X_t = \int_{(-\pi,\pi]} e^{it\lambda}\, dZ(\lambda)$$

$$= \int_{(-\pi,\pi]} [\cos t\lambda + i \sin t\lambda]\, dZ(\lambda)$$

where $Z(\lambda)$ is a random process on $(-\pi,\pi]$

This representation gives X_t in terms of a <u>frequency</u> decomposition

λ is a frequency on the range $(-\pi,\pi]$ but $Z(\lambda)$ is symmetric, so only need to consider

$$0 \leq \lambda \leq \pi.$$

For a time series with dominant <u>high</u> frequency components, $f_x(\lambda)$ is large when λ is near π.

For a time series with dominant <u>low</u> frequency components, $f_x(\lambda)$ is large when λ is near zero.

"High Frequency" - rapid variation

"Low Frequency" - slow variation

NOTE We can reparameterise everything in terms of

$$\lambda^* = \frac{\lambda}{2\pi} \quad \text{so that} \quad -\tfrac{1}{2} < \lambda^* \leq \tfrac{1}{2}$$

i.e rewrite the sdf as

$$f_X^*(\lambda^*) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i 2\pi \lambda^* h} \gamma_X(h)$$

$$\gamma_X(h) = \int_{-\frac{1}{2}}^{1/2} e^{i 2\pi \lambda^* h} f_X^*(\lambda^*) \, d\lambda^*$$

λ is the "angular frequency"

λ^* is the "frequency".

7.2 Spectral Representations for ARMA Processes

Recall that a linear process is the output of a time-invariant linear filter acting on a white noise process.

If $\{X_t\}$ is any process then $\{Y_t\}$ given by

$$Y_t = \sum_{k=-\infty}^{\infty} \psi_k X_{t-k}$$

is a linear filtered version of $\{X_t\}$.

Linear filters are the method via which spectral representations are computed for ARMA processes.

Let $\{X_t\}$ be a stationary time series with mean zero and sdf $f_X(\cdot)$. Let a linear time-invariant filter Ψ be defined by

$$\Psi = \{\psi_j : j = 0, \pm 1, \pm 2, \cdots\}$$

be absolutely summable

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Then $\{X_t\}$ defined by

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

is stationary, zero mean, with sdf

$$f_Y(\lambda) = |\Psi(e^{-i\lambda})|^2 f_X(\lambda)$$

$$= \overline{\Psi(e^{-i\lambda})}\, \Psi(e^{i\lambda})\, f_X(\lambda)$$

$$\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$$

$\Psi(e^{-i\lambda})$ – "TRANSFER FUNCTION"

$|\Psi(e^{-i\lambda})|^2$ – "POWER TRANSFER FUNCTION"

To see this result, recall that

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h+k-j)$$

But

$$\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda)\, d\lambda$$

$$\therefore \gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \int_{-\pi}^{\pi} e^{i(h-j+k)\lambda} f_X(\lambda)\, d\lambda$$

$$= \int_{-\pi}^{\pi} \left(\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda} \right) \left(\sum_{k=-\infty}^{\infty} \psi_k e^{ik\lambda} \right) e^{ih\lambda} f_X(\lambda)\, d\lambda$$

$$= \int_{-\pi}^{\pi} e^{ih\lambda} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda} \right|^2 f_X(\lambda)\, d\lambda$$

By the (Inverse) Fourier Transform

$$f_Y(\lambda) = \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda} \right|^2 f_X(\lambda)$$

$$= \left| \Psi(e^{-i\lambda}) \right|^2 f_X(\lambda)$$

Thus the transfer function $\Psi(e^{i\lambda})$ defines the action of the filter on the process $\{X_t\}$

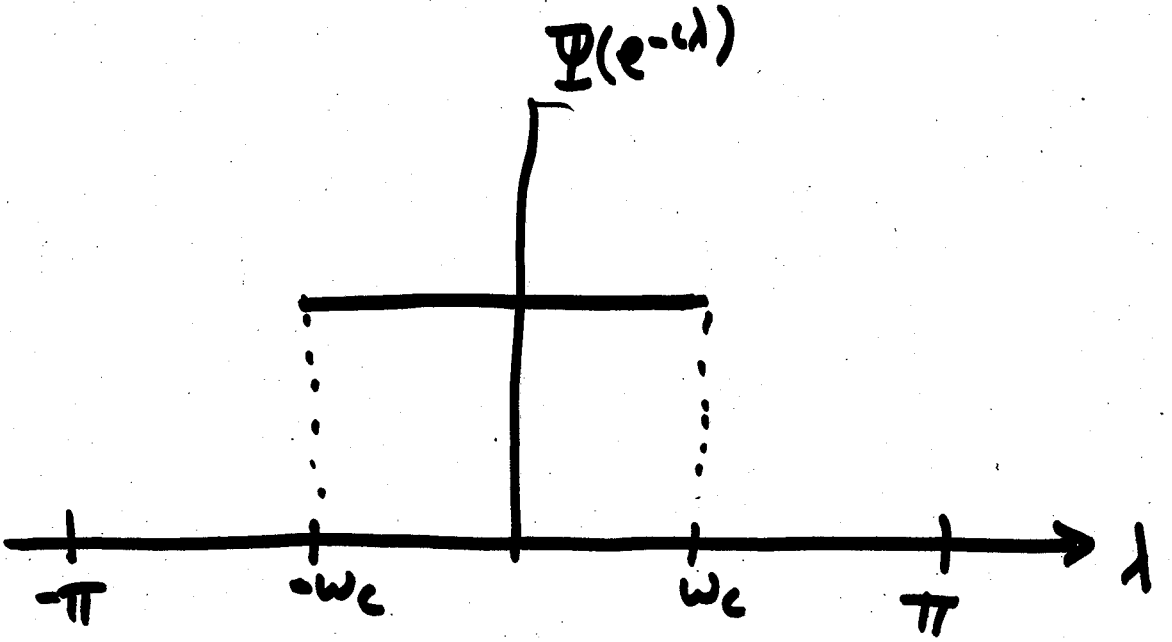
EXAMPLE LOW PASS FILTER

A low pass filter removes high frequency components (similar to noise removal)

Suppose we wish to remove frequency components beyond some threshold ω_c

A filter with this action would have transfer function of the form

$$\begin{cases} 1 & |\lambda| \leq \omega_c \\ 0 & \text{otherwise.} \end{cases}$$



Fourier series expansions can be used to represent "all" functions on bounded intervals

- for discontinuous functions many terms in the series are needed to produce a good approximation.

Consider

$$\psi_j = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{ij\lambda}\,d\lambda = \begin{cases} \frac{\omega_c}{\pi} & j=0 \\ \frac{\sin(j\omega_c)}{j\pi} & j\neq 0 \end{cases}$$

We approximate the "ideal" low pass filter by

$$\Psi_q(e^{-i\lambda}) = \sum_{j=-q}^{q} \psi_j e^{-ij\lambda}$$

for integer q. This approximates the ideal filter as the function

$$\Psi(e^{-i\lambda}) = \begin{cases} 0 & |\lambda| > \omega_c \\ 1 & |\lambda| \leq \omega_c \end{cases}$$

has the Fourier series representation

$$\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$$

EXAMPLE $X_t \sim ARMA(p,q)$

$$\phi(B)X_t = \theta(B)Z_t$$

where $Z_t \sim WN(0,\sigma^2)$

$$f_Z(\lambda) = \frac{\sigma^2}{2\pi}$$

and as $X_t = [\phi(B)]^{-1}\theta(B)Z_t$

By the result for linear filters

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})}\right|^2$$

$$= \frac{\sigma^2}{2\pi}\frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$$

7.3 Model-Free Estimation of the Spectrum

The spectral density function can be estimated from time domain samples.

Let $x_1, \ldots, x_n$ denote the time domain data (for full generality, each x_i is potentially complex-valued).

Define

$$w_k = \frac{2\pi k}{n}$$

for $k = -\left[\frac{n-1}{2}\right], \ldots, \left[\frac{n}{2}\right]$

where $[x]$ is the largest integer less than or equal to x.

$$\left\{\omega_k : k = -\left[\frac{n-1}{2}\right], \ldots \left[\frac{n}{2}\right]\right\} \equiv F_n$$

are known as the FOURIER FREQUENCIES.

Let

$$\underset{\sim}{e}_k = \frac{1}{\sqrt{n}} \begin{bmatrix} e^{i\omega_k} \\ e^{2i\omega_k} \\ \vdots \\ e^{ni\omega_k} \end{bmatrix}$$

Note that

$$\underset{\sim}{\bar{e}}_j^T \underset{\sim}{e}_k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

i.e $\{\underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_n\}$ is an orthonormal basis function for $\mathbb{C}^n$.

i.e $\forall \underset{\sim}{x} \in \mathbb{C}^n$

$$\underset{\sim}{x} = \sum_{k=-[(n-1)/2]}^{[n/2]} a_k \underset{\sim}{e}_k$$

for some coefficients $\{a_k\}$ given by

$$a_k = \bar{\underset{\sim}{e}}_k^T \underset{\sim}{x} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-it\omega_k}$$

$\{a_k\}$ defines the <u>discrete Fourier transform</u>

of $\underset{\sim}{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Note that

$$x_t = \sum_{k=-[(n-1)/2]}^{[n/2]} a_k \left[\cos(\omega_k t) + i \sin(\omega_k t) \right]$$

This gives the decomposition of the Observed sample in a similar form to the Spectral Representation Theorem, discrete version.

- x_t is a linear comporition of sinusoidal functions with frequencies $\omega_k \in F_n$.

DEFINITION THE PERIODOGRAM

The periodogram, $I_n(\lambda)$, is the function

$$I_n(\lambda) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t e^{-it\lambda}\right|^2$$

NOTE $I_n(\omega_k) = |a_k|^2$

NOTE
$$\underset{\sim}{x}^T\underset{\sim}{x} = \sum_{k=-[(n-1)/2]}^{[\frac{n}{2}]} |a_k|^2$$
$$= \sum_{k=-[(n-1)/2]}^{[\frac{n}{2}]} I_n(\omega_k)$$

i.e the "total variation" in $\underset{\sim}{x}$ is the sum of contributions at $\omega_k \in F_n$.

NOTE

$$I_n(\omega_k) = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{it\omega_k}\right)\left(\frac{1}{\sqrt{n}} \sum_{s=1}^{n} x_s e^{-is\omega_k}\right)$$

$$= \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} (x_t x_s) e^{-i(s-t)\omega_k}$$

$$= \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} (x_t - \bar{x})(x_s - \bar{x}) e^{-i(s-t)\omega_k}$$

$$= \sum_{|h|<n} \hat{\gamma}_n(h) e^{-ih\omega_k}$$

$h = s - t.$

as $\sum_{t=1}^{n} e^{-it\omega_k} = 0$ if $\omega_k \neq 0.$

Thus $I_n(\lambda)$ is a sample based estimate of $2\pi f_x(\lambda)$

- unfortunately $I_n(\lambda)$ is not a consistent estimator as, for any

$$0 < \lambda_1 < \ldots < \lambda_m < \pi$$

$$(I_n(\lambda_1), \ldots, I_n(\lambda_m))^T$$

$$\xrightarrow{d} (V_1, \ldots, V_m)^T$$

where $V_j \sim Exp([2\pi f_x(\lambda_j)]^{-1})$

1.e

$$E[I_n(\lambda_j)] \longrightarrow 2\pi f_x(\lambda_j)$$

as $n \to \infty$, so that $I_n(\lambda_j)$ is <u>asymptotically unbiased</u>, but

$$I_n(\lambda_j) \xrightarrow{a.s} \not\longrightarrow 2\pi f_x(\lambda_j)$$

as $n \to \infty$.

To obtain a consistent estimator, we can use a local averaging process to drive the asymptotic variance to zero.

1.e Define

$$\hat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{|j| \le m_n} W_n(j) I_n\left(g(n,\lambda) + 2\pi \frac{j}{n}\right)$$

where "bandwidths" m_n satisfy

$$m_n \longrightarrow \infty \;, \quad \frac{m_n}{n} \longrightarrow 0 \quad \text{as } n \to \infty$$

and "weights" W_n satisfy

$$\forall j : \; W_n(j) = W_n(-j) \;, \; W_n(j) \ge 0$$

$$\sum_{|j| \le m_n} W_n(j) = 1$$

$$\sum_{|j| \le m_n} W_n^2(j) \longrightarrow 0 \quad \text{as } n \to \infty.$$

With these specifications $\hat{f}_n(\lambda)$ is consistent for $f_x(\lambda)$.

EXAMPLE **MOVING AVERAGE**

$$m_n = \sqrt{n}$$

$$W_n(j) = \frac{1}{2m_n+1} \qquad |j| \le m_n$$

NOTE Different choices of $\{m_n, W_n\}$ yield different asymptotic variances for the estimator.

7.4 LIKELIHOOD-BASED INFERENCE FOR PARAMETRIC MODELS

For parametric models (eg ARMA) likelihood inference in the frequency domain can be carried out using a large sample approximation to the (time domain) likelihood known as the

"WHITTLE LIKELIHOOD"

- The "likelihood" is also justified under other criteria - it is a suitable "objective" function.

Recall that for the periodogram,

for $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_m < \pi$

as $n \to \infty$

$$I_n(\lambda_j) \xrightarrow{d} V_j$$

where

$$V_j \sim \text{Exp}\left([2\pi f_x(\lambda_j)]^{-1}\right)$$

i.e $E[V_j] = 2\pi f_x(\lambda_j)$

$$E[I_n(\lambda_j)] \longrightarrow 2\pi f_x(\lambda_j)$$

with $V_1, \ldots, V_j, \ldots, V_m$ <u>independent</u>

i.e For large n, the pdf of V_j (and hence $I_d(\lambda_j)$) is

$$(2\pi f_x(\lambda_j))^{-1} \exp\{-[2\pi f_x(\lambda_j)]^{-1} V_j\}$$

and the log pdf is

$$-\log(2\pi) - \log f_x(\lambda_j) - [2\pi f_x(\lambda_j)]^{-1} V_j$$

This allows us to construct a likelihood once we pick the λ_j

Let

$$\omega_j = \frac{2\pi j}{n}$$

for $j = -\left[\frac{n-1}{2}\right], \ldots, \left[\frac{n}{2}\right]$

denote the Fourier Frequencies on $(-\pi, \pi]$

Suppose that $f_X(\lambda)$ is available pointwise at ω_j, and depends on parameters $\underset{\sim}{\theta}$.

Then the Whittle likelihood is

$$L_W(\theta) = \prod_j \frac{1}{2\pi f_X(\omega_j;\theta)} \exp\left\{-\frac{I_n(\omega_j)}{2\pi f_X(\omega_j;\theta)}\right\}$$

If $\{X_t\} \sim ARMA(p,q)$

$$\phi(B)X_t = \theta(B)Z_t$$

$Z_t \sim WN(0,\sigma^2)$, then

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$$

$$= \frac{\sigma^2}{2\pi} g(\lambda)$$

say. Then the log Whittle likelihood is

$$\log L_W(\theta,\phi) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2$$
$$-\frac{1}{2\sigma^2}\sum_j \frac{I_n(\omega_j)}{g(\omega_j)} - \frac{1}{2}\sum \log g(\omega_j)$$

NOTE A correction to the Whittle likelihood is sometimes used

- if $\omega_j = 0, -\pi, \pi$

then the asymptotic distribution is not correct.

However, not too much information is lost if these Fourier frequencies are omitted.

- The number of terms in the sum is $O(n)$, the amount of information lost is $O(1)$.

DEFINITION

Let $\{X_t\}$ be a stationary process such that for some $\alpha \in (0,1)$ and $c_\rho > 0$ such that

$$\lim_{k \to \infty} \frac{\rho_X(k)}{c_\rho k^{-\alpha}} = 1$$

Then $\{X_t\}$ is a long memory or long range dependence process.

7.5 LONG MEMORY MODELS

Long Memory or Long Range Dependence models are models where the ACVF decays slowly. For ARMA models, the decay in ACF is exponentially fast

i.e $\exists\ r > 1$ s.t

$$r^h \rho_X(h) \longrightarrow 0 \quad \text{as } h \to \infty$$

It is possible to construct similar models where the decay in ACF is much slower: we use

FRACTIONAL INTEGRATION

Recall ARIMA models : for $d = 1, 2, \ldots$

ARIMA(p, d, q) model is

$$(1-B)^d \phi(B) X_t = \theta(B) Z_t$$

Now let d take non-integer values, with

$$-0.5 < d < 0.5$$

We have

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j$$

where

$$\pi_j = \prod_{k=1}^{j} \left(\frac{k-1-d}{k} \right) \quad j = 1, 2, \ldots$$

$$= \binom{d}{j} = \frac{\Gamma(d+1)}{\Gamma(j+1)\,\Gamma(d-j+1)}$$

NOTE Let $W_t = (1-B)^d Z_t$

where $Z_t \sim WN(0, \sigma^2)$, so that

$$\phi(B) X_t = \theta(B) W_t$$

gives an ARMA(p,q) model driven by W_t.

$\{W_t\}$ is **FRACTIONALLY INTEGRATED WHITE NOISE**.

- it is a stationary process with

$$\gamma_W(0) = \frac{\sigma^2 \Gamma(1-2d)}{\{\Gamma(1-d)\}^2}$$

and

$$\rho_W(h) = \frac{\Gamma(h+d)\,\Gamma(1-d)}{\Gamma(h-d+1)\,\Gamma(d)}$$

$$= \prod_{j=1}^{h} \left(\frac{j-1+d}{j-d} \right) \qquad h = 1, 2, \ldots$$

The ACVF of $\{X_t\}$ can be computed using

$$\gamma_X(h) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k \gamma_W(h+j-k)$$

in the usual way.

SPECTRAL DENSITY OF FRACTIONAL ARIMA

Using Transfer function arguments, it is easily seen that if

$$\{X_t\} \sim \text{ARIMA}(p, d, q)$$

for $0 < |d| < \frac{1}{2}$, then

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} |1 - e^{-i\lambda}|^{2d}$$

If $p = q = 0$

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - e^{-i\lambda}|^{2d}}$$

NOTE For ARIMA(0, d, 0)

$$|1 - e^{-i\lambda}| = 2\sin\left(\frac{\lambda}{2}\right)$$

and $$\frac{2\sin\left(\frac{\lambda}{2}\right)}{\lambda} \longrightarrow 1 \quad \text{as } \lambda \to 0$$

$$\therefore \text{ as } \lambda \to 0 \quad f_X(\lambda) \approx \frac{\sigma^2}{2\pi|\lambda|^{2d}}$$

(this holds for general ARMA processes too).

i.e sdf has a pole at $\lambda = 0$.

- This is characteristic of fractionally differenced processes.

FRACTIONAL WHITE NOISE $X_t \sim ARIMA(0, d, 0)$

We have the following infinite MA/AR representations

MA(∞)

$$W_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}$$

where

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \qquad k = 0, 1, \ldots$$

AR(∞)

$$Z_t = \sum_{k=0}^{\infty} \pi_k W_{t-k}$$

$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} \qquad k = 0, 1, \ldots$$

For large k

$$\psi_k \sim \frac{1}{\Gamma(d)} k^{d-1}$$

$$\pi_k \sim \frac{1}{\Gamma(-d)} k^{-d-1}$$

Note For $0 < d < \frac{1}{2}$, if

$$\{X_t\} \sim \text{ARIMA}(p, d, q)$$

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

Then, for large k

$$\gamma_X(k) \sim \text{const} \times |k|^{2d-1}$$