Single Site Streamflow Model 2/5/09

Steps

(1) - Basic exploratory analysis
- mean
- variance / std dev
- skewness
- monthly histograms
- Seasonality in lag 1 correlation
- Seasonality in lag 2 correlation
  (to know then an AR1 model needed. PACF concept) (P 59 + P 64)
- Storage yield plot
  - Think of this as a "statistic" that is an attribute of a time series that you want the simulation to reproduce, that generates storage needed ~~pop~~ capacity
- Correlation function of annual flows (aggregated etc)

(2) Find a Normalizing transform for each month

(3) - Standardize transformed flows

(4) - Look at Seasonal variability of the correlations

(5) - Fit a model.

(6) - Test the whiteness of the residuals

(7) - Test simulations against full suite of statistics and checks

(8) - Consider parameter uncertainty.

Finding a Normalizing transformation

- Lowell's Section 8-4
- Bass p 24 - 26 although this does not do justice to the subject.

The general idea is to find a distribution that fits the marginal distribution, use this to transform data to a normal and estimate correlation of the normally distributed data.

Review Normalization Slide →

May need a standardization step following, depending on transformation

Original data φ

Transformed data $Y = T(\varphi)$

Standardized data $X = \dfrac{Y - \bar{Y}}{\sigma(Y)}$

The X have 0 mean unit variance by construction.

ARMA MODELS - Box Chapter 2   2/10/08

Last time - learned how to transform a nonnormal distribution to Normal.

Today - Focus on the dependence between variables in a time series as quantified by correlation.

The general idea is to construct a correlated sequence from an uncorrelated sequence using a linear filter

Uncorrelated sequence   Correlated sequence

$$a_t \longrightarrow \boxed{\text{FILTER}} \longrightarrow z_t$$

$$\psi(B)$$

For example

$$Z_2 = a_1 + a_2 \qquad E(z_2) = 0$$

$$z_3 = a_3 + a_2 \qquad E(z_3) = 0$$

$$\vdots \qquad \vdots$$

$$\begin{aligned}
\mathrm{Cov}(z_2, z_3) &= E(z_2 z_3) \\
&= E(a_1 a_2 + a_1 a_3 + a_2^2 + a_2 a_3) \\
&= E(a_2^2) \\
&= \sigma_a^2
\end{aligned}$$

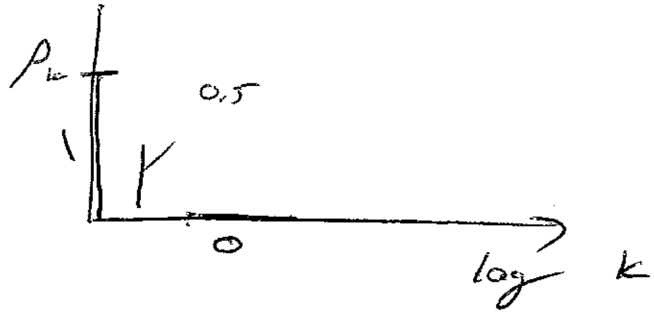
$$\mathrm{Var}(z_2) = E(z_2^2) = E(a_1^2 + 2a_1 a_2 + a_2^2) = 2\sigma_a^2$$

$$\therefore \quad \rho_1 = \frac{\sigma_a^2}{2\sigma_a^2} = 0.5$$

$\cancel{\rho_2 =}$

$$\text{Cov}(z_2, z_4) = \mathbb{E}\big((a_1 + a_2)(a_3 + a_4)\big) = 0$$

ACF (z)



Example in R

These are

$$Z_t = (1 + B)\, a_t$$

where $B a_t = a_{t-1}$ is the backwards shift operator.

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

What if you want higher correlations,

$$Z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

$$= \psi(B)\, a_t$$

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots \qquad \text{(Box 2.8)}$$

$$\psi_0 = 1$$

$$\gamma_k = \text{cov}(z_t, z_{t+k}) = \mathbb{E}\Big((a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots)(a_{t+k} + \psi_1 a_{t+k-1} + \psi_2 a_{t+k-2} + \cdots)\Big)$$

$$= \mathbb{E}\Big(\sum_{j=0}^{\infty}\sum_{h=0}^{\infty} \underbrace{\psi_j a_{t-j} \psi_h a_{t+k-h}}\Big)$$

└ expectation 0 if $j \neq h-k$, $h = j+k$

j	0	1	2	3 ...
h	k	$k+1$	$k+2$	...

$$\therefore \gamma_k = \sum_{j=0}^{\infty} E(a_t^2)\, \psi_j \psi_{j+k}$$

$$= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$$

Specifically $k=0$

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma_z^2$$

$\therefore$ Finite variance requires convergence of this ~~polynomial~~ ~~series~~ series.

~~a condition for this~~

More precisely the infinite polynomial $\psi(B)$ must converge for all B within the unit circle $|B| \leq 1$, treating B as a complex number. (Box p 18)

~~One can also write~~

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

pick coefficients ψ_j to match ρ_k.

This is a <u>moving average process</u>.

This is stationary by construction because PDF of z_t is the same as PDF of z_{t-1}

Now consider a model

$$(1-\phi_1 B)\, z_t = a_t$$

or $\phi(B)\, z_t = a_t$

where $\phi(B) = 1 - \phi_1(B)$

This is

$$z_t - \phi_1 z_{t-1} = a_t$$

or $z_t = \phi_1 z_{t-1} + a_t$

This is an <u>AR(1) model</u>

$$z_{t-1} = \phi_1 z_{t-2} + a_{t-1}$$

$$\therefore\ z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 z_{t-2}$$

$$\uparrow a_{t-2} + \phi z_{t-3}$$

$$= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \cdots$$

$$= (1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots)\, a_t$$

$$= \psi(B)\, a_t$$

We have inverted this AR model into a moving average model.

A general AR model

$$\phi(B)\, z_t = a_t$$

can be inverted to a MA model

$$z_t = \psi(B)\, a_t$$

as long as the polynomial satisfies convergence conditions involving its roots.
Roots of $\phi(B)$ outside unit circle (p 20)

Consider the AR(1) model

$$z_t - \phi_1 z_{t-1} = a_t$$

$$E\left((z_t - \phi_1 z_{t-1})^2\right) = E\left(a_t^2\right)$$

$$\therefore \sigma_z^2 - 2\phi_1 \gamma_1 + \phi_1^2 \sigma_z^2 = \sigma_a^2$$

$$E\left(z_{t-1} z_t\right) = E\left(z_{t-1}\left(a_t + \phi_1 z_{t-1}\right)\right)$$

$$\therefore \gamma_1 = \phi_1 \sigma_z^2$$

$$\therefore \sigma_z^2\left(1 - 2\phi_1^2 + \phi_1^2\right) = \sigma_a^2$$

$$\therefore \sigma_z^2 = \frac{\sigma_a^2}{1-\phi_1^2}$$

$$\phi_1 = \frac{\gamma_1}{\sigma_z^2} = \rho_1$$

$$\therefore \sigma_z^2 = \frac{\sigma_a^2}{1-\rho_1^2}$$

$$\sigma_a^2 = \left(1-\rho_1^2\right)\sigma_z^2$$

This was left for AR(1) model used in HW 2.

$$\gamma_2 = E\left(z_{t-2} z_t\right) = E\left(z_{t-2}\left(a_t + \phi_1 z_{t-1}\right)\right)$$
$$= \phi \gamma_1$$

$\therefore$ Divide by γ_0

$$\rho_2 = \phi \rho_1 = \rho_1^2$$

Similarly $\gamma_3 = \phi_1 \gamma_2$

$$\therefore \rho_3 = \rho_1^3 \quad \text{and} \quad \rho_k = \rho_1^k$$

EXAMPLE IN R

General AR(p) model

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots \phi_p z_{t-p} + a_t$$

$$\gamma_k = E(z_t z_{t-k})$$

$$= E\left(z_{t-k}\left(\phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots \phi_p z_{t-p} + a_t\right)\right)$$

$$= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots \phi_p \gamma_{k-p}$$

Dividing by γ_0

$$\therefore \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \cdots \phi_p \rho_{k-p}$$

or specifically

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots \phi_p \rho_{p-1}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 + \cdots \phi_p \rho_{p-2}$$

$$\vdots$$

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots \phi_p$$

These are the Yule - walker equations and given $\rho_1 \cdots \rho_p$ can be used to solve for ~~$\rho_1 \cdots \rho_p$~~ $\phi_1 \cdots \phi_p$

Box p21 - 34 discusses properties of AR processes in ~~general~~ detail, showing relationships between parameters + ACF.

Combining AR & MA concepts leads to ARMA models

eg $(1 - \phi_1 B) z_t = (1 - \theta_1 B) a_t$

or $\phi(B) z_t = \theta(B) a_t$ ARMA

p40 – 42 relate parameters to correlations

Partial autocorrelation

For a MA process of order q
$ACF(k) = 0$ for $k > q$
because there are no common a_t terms.

We have seen the symmetry between AR & MA processes.

This extends to the concept of Partial autocorrelation – PACF.

If a process is $AR(p)$
and a higher order AR process is fit
ϕ_k $k > p$ should be 0.

PACF is defined as last ϕ in fitting and $AR(k)$ model,
i.e. $\phi_{k,k}$ – from Yule – W equations or Box p59
or R function pacf.

- ACF & PACF are used in identifying a model
ACF truncates for MA
PACF truncates for AR.

General analysis steps

- identification
- parameter estimation - fitting candidates
  - selecting models
- Verification
  - check the independence of residuals

Finish R Example

CE 7430 2/12/07

Last time

General ARMA methodology for simulation of time series to match the correlation structure of observed time series.

$$Z_t = \psi(B)\, a_t \qquad \text{MA}$$

$$\phi(B)\, Z_t = a_t \qquad \text{AR}$$

$$Z_t = \frac{1}{\phi(B)}\, a_t \qquad \text{- invertibility}$$

$$\phi(B)\, Z_t = \psi(B)\, a_t \qquad \text{ARMA } (p, q)$$

p number of AR terms
q " " MA terms

ACF and PACF used for system identification

Figure 2.6 shows steps
- identification
- parameter estimation
- adequacy check / verification

- An iterative process for model development

- Finish the R Example to build intuition for model selection.

- AIC + BIC to avoid over parameterization

- Questions on HW