

Prediction

- BLUP Best Linear Unbiased Predictor

E = expectation = average in population. $E\{X\}=\mu, E\{(X-\mu)^2\}=\sigma^2$

- Predictor (of Y_{n+L})
- Linear ($\hat{Y}_{n+L} = b_n Y_n + b_{n-1} Y_{n-1} + \dots + b_1 Y_1$, assuming means 0)
- Unbiased ($E(Y_{n+L} - \hat{Y}_{n+L}) = 0$)
- Best (pick b_j to minimize something, e.g. $E(Y_{n+L} - \hat{Y}_{n+L})^2$)

$$\begin{matrix} & Y_1 & Y_2 & Y_3 \\ Y_1 & \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} & & \\ Y_2 & & & \\ Y_3 & & & \end{matrix} \quad E \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

$$\hat{Y}_3 = \mu_3 + b_1(Y_1 - \mu_1) + b_2(Y_2 - \mu_2) \quad \text{unbiased}$$

$$\text{Var}\{ Y_3 - \mu_3 - b_1(Y_1 - \mu_1) - b_2(Y_2 - \mu_2) \} =$$

$$\begin{pmatrix} -b_1 & -b_2 & 1 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} \begin{pmatrix} -b_1 \\ -b_2 \\ 1 \end{pmatrix}$$

- Minimize! (set $\frac{\partial}{\partial b_j} (*) = 0$ for $j=1,2$)

$$\text{Solution: } \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}^{-1} \begin{pmatrix} v_{13} \\ v_{23} \end{pmatrix}$$

- Example: $\mu_3 = \mu_2 = \mu_1 = 100, Y_1 = 120, Y_2 = 180$

$$\begin{matrix} & Y_1 & Y_2 & Y_3 \\ Y_1 & \begin{pmatrix} 8 & 4 & 2 \\ 4 & 8 & 4 \\ 2 & 4 & 8 \end{pmatrix} & & \\ Y_2 & & & \\ Y_3 & & & \end{matrix} \quad \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{1}{64-16} \begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 48^{-1} \begin{pmatrix} 0 \\ 24 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

$$\underline{\text{MARKOV}} \quad \hat{Y}_3 = 100 + (1/2)(180-100) = 140$$

(A) Toeplitz covariance matrix: $\text{Cov}(Y_i, Y_j) = \gamma(|i-j|) = \text{fn of } |i-j| \text{ only.}$

$$\begin{pmatrix} A & B & C \\ B & A & B \\ C & B & A \end{pmatrix} \quad \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix}$$

(B) Means all the same

(A)&(B) $\stackrel{\text{defn.}}{< >}$ "COVARIANCE STATIONARY" or just "STATIONARY"

True or false for Toeplitz covariance matrix:

- Covariance matrix of Y_1, Y_2, \dots, Y_n same as that of Y_n, Y_{n-1}, \dots, Y_1
- BLUP of Y_1 based on Y_2, \dots, Y_n uses same weights (b_j) as BLUP of Y_n based on Y_{n-1}, \dots, Y_1
- Stationary: Toeplitz matrix still has n entries, with mean that's $n+1$ parameters to estimate - still too many.
- Idea: Express $\gamma(h)$ as function of just a few unknowns.

Example 1: $\gamma(0), \gamma(1), \gamma(2)$, and $\gamma(h)=0$ if $h>2$. "MA(2)"

Example 2: $\gamma(0)$, and $\gamma(h)=\rho^h\gamma(0)$ for $h>0$. "AR(1)"

Example 3: $\gamma(0), \gamma(1)$, and $\gamma(h)=\rho^h\gamma(1)$ for $h>1$. "ARMA(1,1)"

Example 4: $\gamma(0)$, and $\gamma(h)=0$ if $h>0$. "White Noise"

Check: $\gamma(0) = 100, \gamma(1)=80, \gamma(2)=72, \gamma(3)=64.8, \gamma(4)=58.32$ etc.

What type is this???

- Matrices must be "positive semi-definite" (conditions that prevent negative variances)

Example

$$\begin{pmatrix} 10 & 8 & 0 \\ 8 & 10 & 8 \\ 0 & 8 & 10 \end{pmatrix} \text{ looks like MA(1) but variance of } W = (1, -1, 1) \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

would then be

$$(1 \quad -1 \quad 1) \begin{pmatrix} 10 & 8 & 0 \\ 8 & 10 & 8 \\ 0 & 8 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 30 - 32 = -2 < 0$$

Can't have negative variance!!

- We will see in the course that

(1) All AR, MA, and ARMA models can be expressed in terms of white noise as $Y_{t-\mu} = \alpha_1(Y_{t-1-\mu}) + \alpha_2(Y_{t-2-\mu}) + \dots + \alpha_p(Y_{t-p-\mu}) + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$

(2) Stationarity for these models is ensured if roots of a certain "characteristic polynomial" are in the right region (no unit roots!)

(3) The covariances $\gamma(h)$ can be estimated without assuming a model and thus can serve as identifying functions to show what kind of model is appropriate.

Regression

- Regression may be appropriate for time series
- Time t and seasonal dummies often used
- $Y = X\beta + e$. Example, n=40, trend & quarterly effects

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ \vdots \\ Y_{40} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 1 \\ 1 & 4 & 0 & 0 & 0 \\ 1 & 5 & 1 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 40 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ \vdots \\ e_{40} \end{pmatrix}$$

Quarter 1: $Y_t = \alpha + \beta t + \delta_1 + e_t$

Quarter 2: $Y_t = \alpha + \beta t + \delta_2 + e_t$

Four parallel lines

Quarter 3: $Y_t = \alpha + \beta t + \delta_3 + e_t$

Quarter 4: $Y_t = \alpha + \beta t + 0 + e_t$

- $\hat{\beta} = (X'X)^{-1}(X'Y)$ B.L.U.E. if errors are iid
- $(X'X)^{-1}(MSE)$ is proper variance-covariance matrix if errors are iid
- if errors are not iid then:
 - $\hat{\beta} = (X'X)^{-1}(X'Y)$ unbiased but not best
 - $(X'X)^{-1}(MSE)$ not appropriate
 - t tests, P-values, F tests all wrong (they use $(X'X)^{-1}(MSE)$)

- How to tell?

Durbin-Watson test:

Run regression as usual (Ordinary Least Squares, OLS)

Get residuals r_t . Compute $D = \frac{\sum_{t=2}^n (r_t - r_{t-1})^2}{\sum_{t=1}^n r_t^2}$

For i.i.d. r_t you'd have $E\{\sum_{t=2}^n (r_t - r_{t-1})^2\} = 2(n-1)\sigma^2$ and $E\{\sum_{t=1}^n r_t^2\} = n\sigma^2$

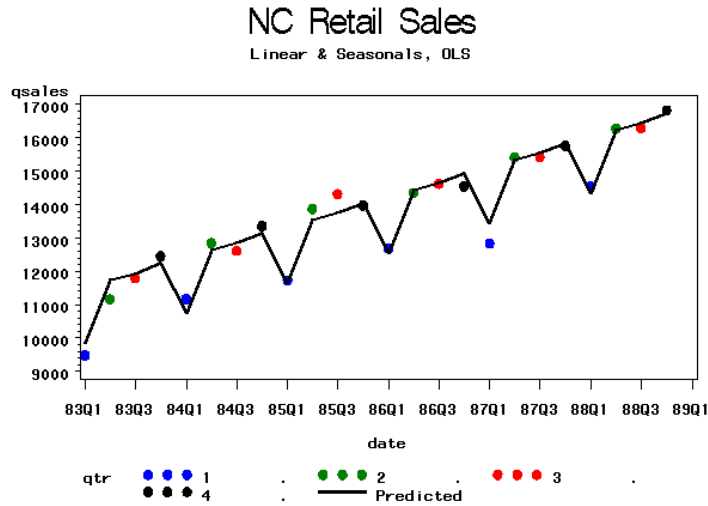
For i.i.d. D should be near 2.

If r_t and r_{t-1} alike (positively correlated) $\sum_{t=2}^n (r_t - r_{t-1})^2$ smaller and $D < 2$.

Durbin-Watson give bounds on critical value and computationally intensive method to get exact p-values.

- Example: Quarterly NC retail sales.

```
options ls=76;
title 'North Carolina Retail Sales in million $';
title2 "Quarterly starting in 1983";
Data NCSALES;
input qsales t t2 s1 s2 s3 s4 date :yyq6.;
qtr=qtr(date); x=t+.3; *(for graphs);
cards;
          9485.68      1      1      1      0      0      0      1983Q1
          11164.09     2      4      0      1      0      0      1983Q2
                (more data)
          16829.22     24     576     0      0      0      1      1988Q4
;
data next; set ncsales;
proc reg;
model qsales = t S1 S2 S3/dw;
output out=out1 p=pred;
* Autoreg shows Durbin-Watson 1.2190    Pr < DW 0.0289;
proc plot; plot  qsales*t=qtr pred*x="+"/overlay vpos=26;
run;
```



The REG Procedure
Model: MODEL1
Dependent Variable: qsales

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 4 | 78085176 | 19521294 | 189.90 | <.0001 |
| Error | 19 | 1953171 | 102798 | | |
| Corrected Total | 23 | 80038346 | | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1 | 11342 | 187.41080 | 60.52 | <.0001 |
| t | 1 | 224.91923 | 9.58041 | 23.48 | <.0001 |
| s1 | 1 | -1737.00230 | 187.32915 | -9.27 | <.0001 |
| s2 | 1 | -49.68154 | 186.10022 | -0.27 | 0.7924 |
| s3 | 1 | -82.12577 | 185.35894 | -0.44 | 0.6627 |

Durbin-Watson D 1.219
Number of Observations 24
1st Order Autocorrelation 0.358

* Autoreg shows Durbin-Watson 1.2190 Pr < DW 0.0289;

- To predict, concatenate future values of all X's to data (set Y = missing (.))
- Advantages:
 - Easy to understand and implement
 - Picks up very regular trends and seasonal patterns (dummy variables)
 - Has check for autocorrelation (Durbin_Watson)
- Disadvantages
 - Not flexible (changing trend, seasonal difficult to model)
 - Autocorrelation destroys inference (but see PROC AUTOREG later)
 - Need future values of input variables.

• **Example:** Visits to Toronto hospital for asthma related symptoms.

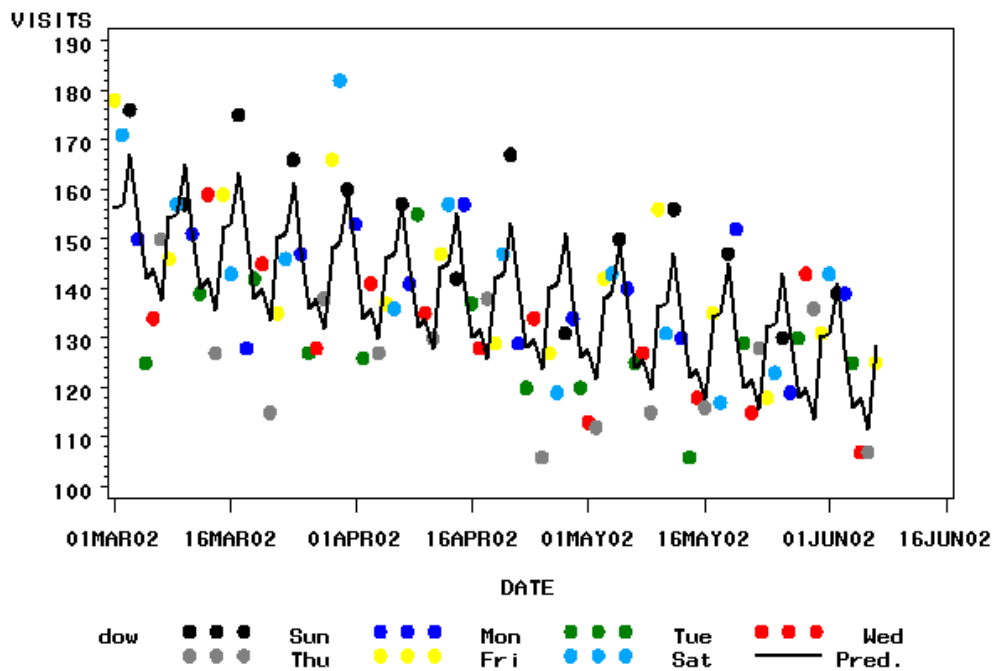
```
proc reg data=Asthma;
model visits = date d1-d6/dw;
output out=out1 predicted=p; run;
proc plot data=out1;
plot visits*date=dow
      p*date=dow/vpos=30;
```

d1 through d6 are weekday indicators (d1=Sunday)

| Analysis of Variance | | | | | |
|----------------------|----|----------------|-------------|---------|--------|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 7 | 15349 | 2192.70743 | 15.31 | <.0001 |
| Error | 91 | 13034 | 143.22830 | | |
| Corrected Total | 98 | 28383 | | | |

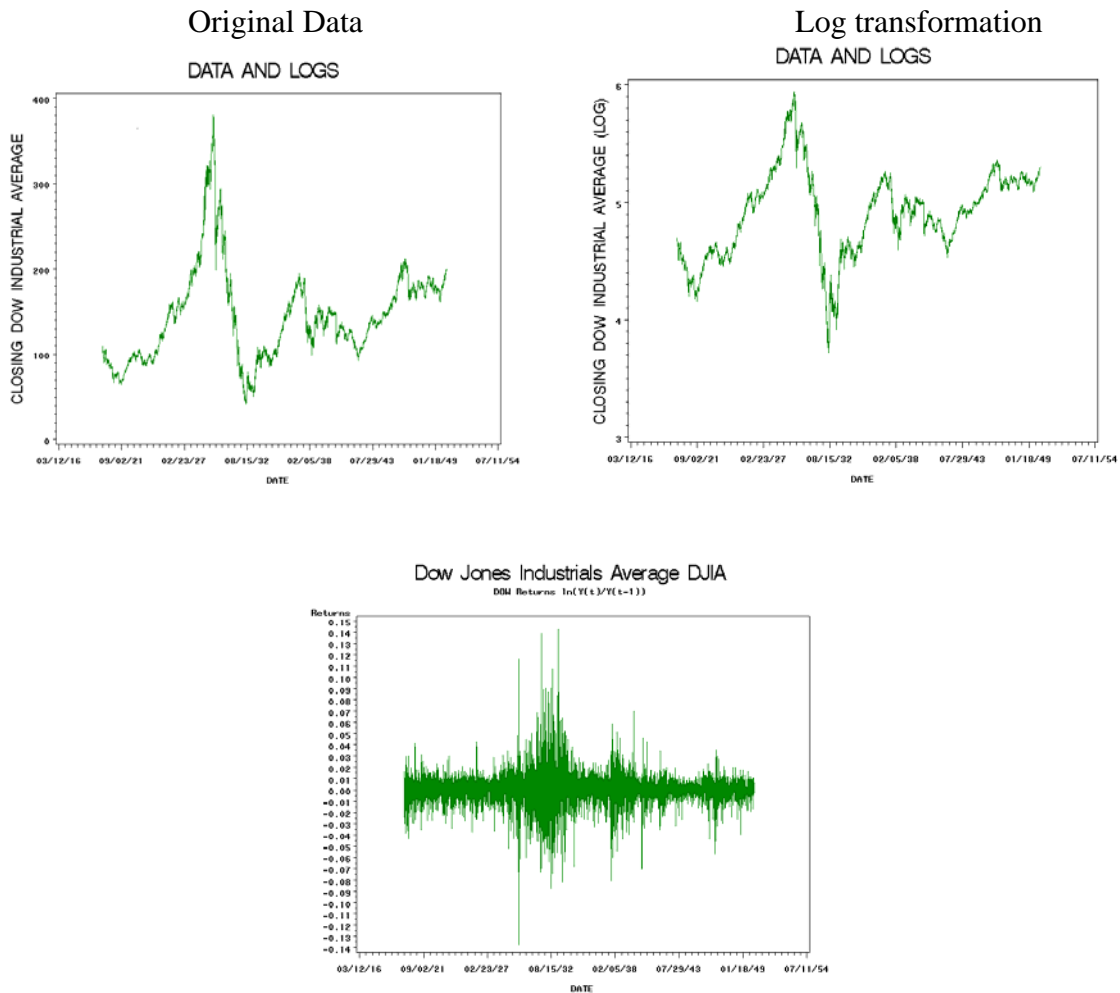
| Parameter Estimates | | | | | |
|---------------------|----|--------------------|----------------|---------|---------|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
| Intercept | 1 | 4563.92141 | 651.13157 | 7.01 | <.0001 |
| DATE | 1 | -0.28615 | 0.04215 | -6.79 | <.0001 |
| D1 | 1 | 10.14329 | 4.52360 | 2.24 | 0.0274 |
| D2 | 1 | -2.64199 | 4.52419 | -0.58 | 0.5607 |
| D3 | 1 | -14.07013 | 4.52517 | -3.11 | 0.0025 |
| D4 | 1 | -12.28398 | 4.52655 | -2.71 | 0.0080 |
| D5 | 1 | -17.85497 | 4.52831 | -3.94 | 0.0002 |
| D6 | 1 | -1.14653 | 4.44862 | -0.26 | 0.7972 |

Hospital elective visits



Transformations

- Most common: no transformation or log
Try log if variation increases as mean increases.
- Example: plot of Dow Jones, Log(Dow), Log difference
Transformation can have big effect.



- Log difference (above) often used in economics (Log is natural log, ln)

$$\text{Log}(Y_t) - \text{Log}(Y_{t-1}) = \text{Log}(Y_t/Y_{t-1})$$

Taylor's series: $\text{Log}(1+\epsilon) = \log(1) + 1\epsilon - \epsilon^2/2 + \dots \approx \epsilon$ for ϵ small.

Thus $100 \text{Log}(Y_t/Y_{t-1})$ is approximate percentage change if small.

```

data a; array X(11);
do i=1 to 11; X(i) = .88+i/50; end; output;
do i=1 to 11; x(i) =log(x(i)); end; output;
proc print noobs; var X1-X11; format X1-X11 5.2;
run;

```

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 |
|-------|-------|-------|-------|-------|------|------|------|------|------|------|
| 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 |
| -0.11 | -0.08 | -0.06 | -0.04 | -0.02 | 0.00 | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 |

• $\log(Y)=X \sim N(\mu, \sigma^2) \Rightarrow E\{Y\} = E\{e^X\} = \frac{1}{\sqrt{2\pi}\sigma} \int e^X e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx =$
 $\frac{1}{\sqrt{2\pi}\sigma} \int e^{-\frac{X^2-2(\mu+\sigma^2)X+\mu^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} e^{\mu+\sigma^2/2} \int e^{-\frac{X^2-2(\mu+\sigma^2)X+(\mu+\sigma^2)^2}{2\sigma^2}} dx =$
 $e^{\mu+\sigma^2/2}$ (not just e^μ)

• Although exponentiating mean of $\log(Y)$ does not give mean of Y , it is true that $\Pr\{\log(Y) < C\} = \Pr\{Y < e^C\}$ from which we see that exponentiating median of $\log(Y)$ gives the median of the original values.

White noise: e_t mean 0, variance σ^2 uncorrelated

Moving Average

Order 1: $(Y_t - \mu) = e_t - \theta_1 e_{t-1}$ all t "MA(1)"

Order q: $(Y_t - \mu) = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$ all t "MA(q)"

Autoregressive - AR(1)

$(Y_t - \mu) = \rho(Y_{t-1} - \mu) + e_t$ all t

$E(Y_t) - \mu = \rho[E(Y_{t-1}) - \mu]$

Stationarity:

$E(Y_t)$ constant all t (call it μ)

$Cov(Y_t, Y_{t-j}) = \gamma(j) =$ function of j only

$E(Y_t) = \mu$ (if $|\rho| < 1$)

Assuming $|\rho| < 1$

$(y_t - \mu) = \rho(Y_{t-1} - \mu) + e_t$
 $= \rho[\rho(Y_{t-2} - \mu) + e_{t-1}] + e_t$
 $= \rho[\rho[\rho(Y_{t-3} - \mu) + e_{t-2}] + e_{t-1}] + e_t$
 etc.

$(Y_t - \mu) = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \dots$

$E(Y_t) = \mu$

$Var(Y_t) = \sigma^2(1 + \rho^2 + \rho^4 + \rho^6 + \dots) = \sigma^2 / (1 - \rho^2)$

$Cov(Y_t, Y_{t-j}) = \rho^j \sigma^2 / (1 - \rho^2)$

Example:

Plot of $\gamma(j)$ versus j is Autocovariance Function

| | | | | | | | | |
|-------------|----|----|----|---|---|---|---|-----|
| $j=$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $\gamma(j)$ | 64 | 32 | 16 | 8 | 4 | 2 | 1 | 1/2 |

(1) Find ρ , variance of Y_t variance of e_t and μ :

$\rho = 0.5$ (geometric decay rate)

Variance of Y_t is 64.

Variance of e : [Using $\gamma(0) = 64 = \sigma^2/(1-\rho^2) = \sigma^2/(1-.25)$]

$\sigma^2 = 64(.75) = 48.$

Covariances have *no* information about μ .

Forecast:

$Y_{n+1} = \mu + \rho(Y_n - \mu) + e_{n+1}$

$\mu = 90$ known (or estimated)

Data Y_1, Y_2, \dots, Y_n with $Y_n = 106$. We see that $\rho = 0.5$

$\hat{Y}_{n+1} = 90 + .5(106 - 90) = 98$ error $Y_{n+1} - \hat{Y}_{n+1} = e_{n+1}$

$Y_{n+2} = \mu + \rho^2(Y_n - \mu) + e_{n+2} + \rho e_{n+1}$

$\hat{Y}_{n+2} = \mu + \rho^2(Y_n - \mu) = 94$, error = $e_{n+2} + \rho e_{n+1}$

$\hat{Y}_{n+j} = \mu + \rho^j(Y_n - \mu)$ error = $e_{n+j} + \rho e_{n+j-1} + \dots + \rho^{j-1} e_{n+1}$

98, 94, 92, 91, 90.5, \dots forecasts.

Forecast intervals (large n).

μ, ρ known (or estimated and assumed known)

Forecast errors $e_{n+1}, e_{n+2} + \rho e_{n+1}, e_{n+3} + \rho e_{n+2} + \rho^2 e_{n+1}, \dots$

(1) Can't know future e's

(2) Can estimate variance $\sigma^2(1 + \rho^2 + \dots + \rho^{2j-2})$.

(3) Estimate σ^2 : Use $r_t = (Y_t - \mu) - \rho(Y_{t-1} - \mu)$ then

$\hat{\sigma}^2 = \Sigma r_t^2 / n$ or

Get $S_y^2 = \Sigma(Y_t - \bar{Y})^2 / n$ then $\hat{\sigma}^2 = S_y^2(1 - \rho^2)$

$$\hat{Y}_{n+j} \pm 1.96 \sqrt{\hat{\sigma}^2(1 + \rho^2 + \dots + \rho^{2j-2})}$$

Estimating an AR(1)

$$Y_t = \mu(1-\rho) + \rho Y_{t-1} + e_t = \lambda + \rho Y_{t-1} + e_t$$

Looks like a regression: Regress Y_t on 1, Y_{t-1} or $Y_t - \bar{Y}$ on $Y_{t-1} - \bar{Y}$ (no int)

1. $n^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{Y})^2$ converges to $E\{ (Y_t - \mu)^2 \} = \gamma(0) = \frac{\sigma^2}{1-\rho^2}$

2. $\sqrt{n} [\sum (Y_{t-1} - \bar{Y}) e_t / n]$ is \sqrt{n} times a mean of $(Y_{t-1} - \mu) e_t$ terms *uncorrelated* (but not independent)

Nevertheless $\sqrt{n} [\sum (Y_{t-1} - \bar{Y}) e_t / n]$ converges to $N(0, ?)$ where variance is $E\{ (Y_{t-1} - \mu)^2 e_t^2 \} = E\{ (Y_{t-1} - \mu)^2 \} E\{ e_t^2 \} = \gamma(0) \sigma^2$

3. $\sqrt{n} (\hat{\rho} - \rho) = \sqrt{n} [\sum (Y_{t-1} - \bar{Y}) e_t / n] / [n^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{Y})^2]$

in the limit this is $N(0, \gamma(0) \sigma^2 / \gamma(0)^2) = N(0, 1-\rho^2)$.

Estimating MA(q)

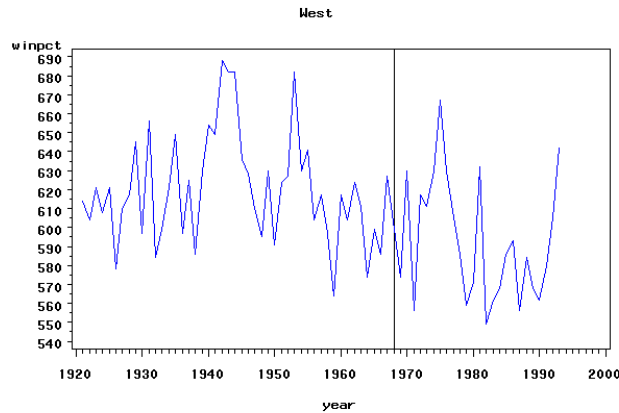
Least squares:

Compute $e(t) = (Y_t - \mu) + \theta_1 e(t-1) + \dots + \theta_q e(t-q)$
 Search over parameters μ, θ_j to minimize $\sum_{t=1}^n e^2(t)$

Alternatively, maximize the likelihood for AR, MA, or ARMA.

EXAMPLE:

Winning Percentage (x 10) for baseball's National League pennant winner.



Regression:

| Year | Y_t | | Y_{t-1} |
|------|-------|---|-----------|
| 1921 | 614 | 1 | . |
| 1922 | 604 | 1 | 614 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 1993 | 642 | 1 | 605 |

$$\text{PROC REG: } \hat{Y}_t = 341.24 + .44116 Y_{t-1} + e_t, \quad s^2 = 863.7$$

(66.06) (.1082)

$$\hat{Y}_t - 610.62 = .44116 (Y_{t-1} - 610.62) + e_t$$

| Year | Forecast | Forecast Standard Error |
|-------|----------------------------------|---------------------------------|
| 1994: | $341.24 + .44 (642) = 624.46$ | $\sqrt{863.7} = 29.4$ |
| 1995: | $341.24 + .44 (624.46) = 616.73$ | $\sqrt{863.7(1+.44^2)} = 32.10$ |

or

$$1995: 610.62 + .44^2(642-610.62)$$

$$2054: 610.62 + .44^{60}(41.38) = 610.62 = \hat{\mu} \quad \sqrt{863.7/(1-.44^2)} = \sqrt{\hat{\sigma}^2(0)}$$

so long term forecast is just mean.

Theory assigns std. error $\sqrt{(1-\rho^2)/n}$ to $\hat{\rho}$. We have $\sqrt{(1-.44^2)/73} = .105$

Identification - Part 1

Auto correlation $\rho(j) = \gamma(j) / \gamma(0)$
 (For AR(1) , $\rho(j) = \rho^j$) ACF

Partial Autocorrelation

Regress Y_t on $Y_{t-1}, Y_{t-2}, \dots, Y_{t-j}$

Last coefficient Π_j is called j^{th} partial autocorrelation PACF.

More formally, $\hat{\gamma}(j) = n^{-1} \sum (Y_t - \mu)(Y_{t-j} - \mu)$ estimates $\gamma(j)$

$\frac{1}{n} \mathbf{X}'\mathbf{X}$ regression matrix looks like

$$\frac{1}{n} \begin{bmatrix} \Sigma(Y_{t-1} - \bar{Y})^2 & \Sigma(Y_{t-1} - \bar{Y})(Y_{t-2} - \bar{Y}) & \dots \\ \Sigma(Y_{t-1} - \bar{Y})(Y_{t-2} - \bar{Y}) & \Sigma(Y_{t-2} - \bar{Y})^2 & \dots \\ \vdots & & \ddots \end{bmatrix}$$

so formally, $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$ is analogous to the population equation (also "best predictor" idea)

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(j-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(j-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(j-1) & \gamma(j-2) & \dots & \gamma(0) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_j (= \Pi_j) \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(j) \end{bmatrix}$$

This defines $\Pi_j = j^{\text{th}}$ partial autocorrelation

For AR(1), partials are

| | | | |
|----------------|---|---|-----|
| $j = 1$ | 2 | 3 | ... |
| $\Pi_j = \rho$ | 0 | 0 | ... |

Yule-Walker equations

$$(Y_{t-j} - \mu)[(Y_t - \mu) - .5(Y_{t-1} - \mu)] = (Y_{t-j} - \mu)(e_t + .8e_{t-1})$$

Take expected value

| | |
|------------------------------------------------------------|-----------------------------------------------------------------------|
| $j = 0 \quad \gamma(0) - .5\gamma(1) = \sigma^2(1 + 1.04)$ | $\gamma(0) - \alpha\gamma(1) = \sigma^2(1 - \theta(\alpha - \theta))$ |
| $j = 1 \quad \gamma(1) - .5\gamma(0) = \sigma^2(.8)$ | $\gamma(1) - \alpha\gamma(0) = -\theta\sigma^2$ |
| $j > 1 \quad \gamma(j) - .5\gamma(j-1) = 0$ | |

$$\begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} = \begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix}^{-1} \begin{pmatrix} 2.04 \\ 0.80 \end{pmatrix} \sigma^2 = \begin{pmatrix} 3.2533 \\ 2.4266 \end{pmatrix} \sigma^2$$

| | | | | | | |
|-----------|---|------|------|------|------|------|
| j | 0 | 1 | 2 | 3 | 4 | |
| $\rho(j)$ | 1 | .746 | .373 | .186 | .093 | etc. |

Define $\gamma(-j) = \gamma(j)$, $\rho(-j) = \rho(j)$.

- In general Yule-Walker relates covariances to parameters. Two uses:
- (1) Given model, get $\gamma(j)$ and $\rho(j)$
 - (2) Given estimates of $\gamma(j)$ get rough estimates of parameters.

Identification - Part II

Inverse Autocorrelation IACF

For the model

$$(Y_t - \mu) - \alpha_1(y_{t-1} - \mu) - \dots - \alpha_p(Y_{t-p} - \mu) = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

define IACF as ACF of the dual model:

$$(Y_t - \mu) - \theta_1(Y_{t-1} - \mu) - \dots - \theta_q(Y_{t-q} - \mu) = e_t - \alpha_1 e_{t-1} - \dots - \alpha_p e_{t-p}$$

IACF of AR(p) is ACF of a MA(p)

IACF of MA(q) is ACF of an AR(q)

How do we estimate ACF, IACF, PACF from data?

Autocovariances $\hat{\gamma}(j) = \frac{1}{n-j} \sum_{t=1}^{n-j} (Y_t - \bar{Y})(Y_{t+j} - \bar{Y})/n$

ACF $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$

PACF plug $\hat{\gamma}(j)$ into formal defining formula and solve for $\hat{\Pi}_j$.

IACF: Approximate by fitting long autoregression

$$(Y_t - \mu) = \hat{\alpha}_1(Y_{t-1} - \mu) + \dots + \hat{\alpha}_k(Y_{t-k} - \mu) + e_t$$

Compute ACF of dual model

$$Y_t - \mu = e_t - \hat{\alpha}_1 e_{t-1} - \dots - \hat{\alpha}_k e_{t-k}.$$

To fit the long autoregressive plug $\hat{\gamma}(j)$ into Yule-Walker equations for AR(k), or just regress Y_j on $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$.

All 3 functions IACF, PACF, ACF computed in PROC ARIMA. How do we interpret them? Compare to catalog of theoretical IACF, PACF, ACF for AR, MA, and ARMA models. See SAS System for Forecasting Time Series book for several examples - section 3.3.2.

Variance for IACF, PACF approximately $\frac{1}{n}$

For ACF, SAS uses Bartlett's formula. For $\hat{\rho}(j)$ this is

$$n^{-1} \sum_{i=-j+1}^{j-1} (\hat{\rho}(i))^2$$

(Fuller gives Bartlett's formula as 6.2.11 after first deriving a more accurate estimate of the variance of $\hat{\rho}(i)$. The sum there is infinite so in SAS the hypothesis being tested is $H_0: \rho(j)=0$ assuming $\rho(i)=0$ for $i>j$. Assuming a MA of order no more than j , is the j^{th} autocorrelation 0?)

Syntax

```
PROC ARIMA;
  IDENTIFY VAR=Y (NOPRINT NLAG=10 CENTER);
  ESTIMATE P=2 Q=1 (NOCONSTANT NOPRINT ML PLOT);
  FORECAST LEAD=7 OUT=OUT1 ID=DATE INTERVAL=MONTH;
```

- (1) I, E, F will work.
- (2) Must have I preceding E, E preceding F
- (3) CENTER subtracts \bar{Y}
- (4) NOCONSTANT is like NOINT in PROC REG

- (5) ML (maximum likelihood) takes more time but has slightly better accuracy than the default least squares.
- (6) PLOT gives ACF, PACF, IACF of residuals

Diagnostics: Box-Ljung chi-square on data Y_t or residuals \hat{e}_t .

- (1) Compute estimated ACF $\hat{\rho}(j)$
 - (2) Test is $Q = n(n+2) \sum_{j=1}^k \frac{(\hat{\rho}(j))^2}{(n-j)}$
 - (3) Compare to χ^2 distribution with $k - p - q$ d.f. (ARMA(p, q))
- SAS (PROC ARIMA) will give Q test on original data and on residuals from fitted models.
 - Q statistics given in sets of 6, i.e. for j=1 to 6, for j=1 to 12, for j=1 to 18, etc.
Note that these are cumulative
 - For original series H_0 : Series is white noise to start with.
 - For residuals H_0 : Residual series is white noise.

Suppose residuals autocorrelated - what does it mean?

Can predict future residuals from past - then why not do it?

Model predicts using correlation.

Autocorrelated residuals => model has not captured all the predictability in the data.

So.... H_0 : Model is sufficient vs. H_1 : Needs more work <=> "lack of fit" test

Let's try some examples. All have this kind of header, all have 1500 obs.

```

ARIMA Procedure

Name of variable = Y1.

Mean of working series = -0.03206

Standard deviation      = 1.726685

Number of observations =      1500

```

Y1 Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|--|
| 0 | 2.98144 | 1.00000 | | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 2.39661 | 0.80384 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | 1.89578 | 0.63586 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 3 | 1.49191 | 0.50040 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 4 | 1.20474 | 0.40408 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 5 | 1.00738 | 0.33788 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 6 | 0.8373 | 0.28084 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 7 | 0.67985 | 0.22803 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 8 | 0.58866 | 0.19744 | | | | | | | | | | | . | | **** | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.50067 | | | | | | ***** | | | | | | . | | | | | | | | | | |
| 2 | -0.00386 | | | | | | | | | | | . | | . | | | | | | | | | |
| 3 | 0.02033 | | | | | | | | | | | . | | . | | | | | | | | | |
| 4 | 0.01656 | | | | | | | | | | | . | | . | | | | | | | | | |
| 5 | -0.01834 | | | | | | | | | | | . | | . | | | | | | | | | |
| 6 | -0.02593 | | | | | | | | | | | * | | . | | | | | | | | | |
| 7 | 0.04455 | | | | | | | | | | | . | | * | | | | | | | | | |
| 8 | -0.02228 | | | | | | | | | | | . | | . | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|--|
| 1 | 0.80384 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | -0.02912 | | | | | | | | | | | * | | . | | | | | | | | | |
| 3 | -0.00625 | | | | | | | | | | | . | | . | | | | | | | | | |
| 4 | 0.02961 | | | | | | | | | | | . | | * | | | | | | | | | |
| 5 | 0.03108 | | | | | | | | | | | . | | * | | | | | | | | | |
| 6 | -0.00511 | | | | | | | | | | | . | | . | | | | | | | | | |
| 7 | -0.01304 | | | | | | | | | | | . | | . | | | | | | | | | |
| 8 | 0.03765 | | | | | | | | | | | . | | * | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | |
|-----|---------|------------------|-------|-------|-------|-------|-------|-------|-------|
| Lag | Square | DF | Prob | | | | | | |
| 6 | 2493.02 | 6 | 0.000 | 0.804 | 0.636 | 0.500 | 0.404 | 0.338 | 0.281 |

Y2

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 |
|-----|---------|---------|----|-------|---|---|---|-------|---|---|---|-------|---|-------|---|---|---|---|---|---|---|---|---|
| 0 | 2.84007 | 1.00000 | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | -2.2489 | -.79184 | | ***** | | | | | | | | | | . | | | | | | | | | |
| 2 | 1.80586 | 0.63585 | | | | | | | | | | . | | ***** | | | | | | | | | |
| 3 | -1.4603 | -.51416 | | | | | | ***** | | | | | | . | | | | | | | | | |
| 4 | 1.14644 | 0.40367 | | | | | | | | | | . | | ***** | | | | | | | | | |
| 5 | -0.9219 | -.32460 | | | | | | | | | | ***** | | . | | | | | | | | | |
| 6 | 0.76776 | 0.27033 | | | | | | | | | | . | | **** | | | | | | | | | |
| 7 | -0.6261 | -.22044 | | | | | | | | | | **** | | . | | | | | | | | | |
| 8 | 0.48619 | 0.17119 | | | | | | | | | | . | | *** | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|--|
| 1 | 0.47434 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | 0.00302 | | | | | | | | | | | . | | . | | | | | | | | | |
| 3 | 0.04342 | | | | | | | | | | | . | | * | | | | | | | | | |
| 4 | 0.03132 | | | | | | | | | | | . | | * | | | | | | | | | |
| 5 | -0.01538 | | | | | | | | | | | . | | . | | | | | | | | | |
| 6 | -0.01612 | | | | | | | | | | | . | | . | | | | | | | | | |
| 7 | 0.01805 | | | | | | | | | | | . | | . | | | | | | | | | |
| 8 | 0.01532 | | | | | | | | | | | . | | . | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.79184 | | ***** | | | | | | | | | . | | | | | | | | | | | |
| 2 | 0.02372 | | | | | | | | | | | . | | . | | | | | | | | | |
| 3 | -0.01028 | | | | | | | | | | | . | | . | | | | | | | | | |
| 4 | -0.03180 | | | | | | | | | | | * | | . | | | | | | | | | |
| 5 | -0.01921 | | | | | | | | | | | . | | . | | | | | | | | | |
| 6 | 0.02570 | | | | | | | | | | | . | | * | | | | | | | | | |
| 7 | 0.00905 | | | | | | | | | | | . | | . | | | | | | | | | |
| 8 | -0.02456 | | | | | | | | | | | . | | . | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | | |
|-----|---------|------------------|-------|--------|-------|--------|-------|--------|-------|--|--|--|
| Lag | Square | DF | Prob | | | | | | | | | |
| 6 | 2462.73 | 6 | 0.000 | -0.792 | 0.636 | -0.514 | 0.404 | -0.325 | 0.270 | | | |

Y3

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|--|
| 0 | 1.68768 | 1.00000 | | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 0.87193 | 0.51664 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | 0.92573 | 0.54852 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 3 | 0.60333 | 0.35749 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 4 | 0.54891 | 0.32524 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 5 | 0.43268 | 0.25637 | | | | | | | | | | | . | | **** | | | | | | | | | |
| 6 | 0.38316 | 0.22704 | | | | | | | | | | | . | | **** | | | | | | | | | |
| 7 | 0.28252 | 0.16740 | | | | | | | | | | | . | | *** | | | | | | | | | |
| 8 | 0.26912 | 0.15946 | | | | | | | | | | | . | | *** | | | | | | | | | |

"," marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.17153 | | | | | | | | | | | *** | | . | | | | | | | | | |
| 2 | -0.32421 | | | | | | | | | | | ***** | | . | | | | | | | | | |
| 3 | 0.04133 | | | | | | | | | | | . | | * | | | | | | | | | |
| 4 | 0.02702 | | | | | | | | | | | . | | * | | | | | | | | | |
| 5 | -0.03447 | | | | | | | | | | | * | | . | | | | | | | | | |
| 6 | -0.02051 | | | | | | | | | | | . | | . | | | | | | | | | |
| 7 | 0.03163 | | | | | | | | | | | . | | * | | | | | | | | | |
| 8 | -0.01685 | | | | | | | | | | | . | | . | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|--|
| 1 | 0.51664 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | 0.38414 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 3 | -0.02480 | | | | | | | | | | | . | | . | | | | | | | | | |
| 4 | 0.00940 | | | | | | | | | | | . | | . | | | | | | | | | |
| 5 | 0.03415 | | | | | | | | | | | . | | * | | | | | | | | | |
| 6 | 0.02322 | | | | | | | | | | | . | | . | | | | | | | | | |
| 7 | -0.02602 | | | | | | | | | | | * | | . | | | | | | | | | |
| 8 | 0.02131 | | | | | | | | | | | . | | . | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | | |
|-----|---------|------------------|-------|-------|-------|-------|-------|-------|-------|--|--|--|
| Lag | Square | DF | Prob | | | | | | | | | |
| 6 | 1382.13 | 6 | 0.000 | 0.517 | 0.549 | 0.357 | 0.325 | 0.256 | 0.227 | | | |

Y4

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|-------|-----|---|-----|-------|-------|---|---|---|---|---|---|---|---|--|
| 0 | 1.87853 | 1.00000 | | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 0.90481 | 0.48166 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 2 | -0.3135 | -.16687 | | | | | | | | | *** | | . | | | | | | | | | | | |
| 3 | -0.7114 | -.37872 | | | | | | | | ***** | | . | | | | | | | | | | | | |
| 4 | -0.3603 | -.19181 | | | | | | | | **** | | . | | | | | | | | | | | | |
| 5 | 0.10377 | 0.05524 | | | | | | | | | . | | * | | | | | | | | | | | |
| 6 | 0.24624 | 0.13108 | | | | | | | | | . | | *** | | | | | | | | | | | |
| 7 | 0.13762 | 0.07326 | | | | | | | | | . | | * | | | | | | | | | | | |
| 8 | 0.05574 | 0.02967 | | | | | | | | | . | | * | | | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|-------|---|---|---|---|------|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.60608 | | | | | | ***** | | . | | | | | | | | | | | | | | |
| 2 | 0.27383 | | | | | | | | | . | | **** | | | | | | | | | | | |
| 3 | 0.00795 | | | | | | | | | . | | . | | | | | | | | | | | |
| 4 | 0.00599 | | | | | | | | | . | | . | | | | | | | | | | | |
| 5 | -0.00347 | | | | | | | | | . | | . | | | | | | | | | | | |
| 6 | -0.02802 | | | | | | | | | * | | . | | | | | | | | | | | |
| 7 | 0.03922 | | | | | | | | | . | | * | | | | | | | | | | | |
| 8 | -0.03526 | | | | | | | | | * | | . | | | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|-------|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|---|--|
| 1 | 0.48166 | | | | | | | | | . | | ***** | | | | | | | | | | | |
| 2 | -0.51936 | | | | | | ***** | | . | | | | | | | | | | | | | | |
| 3 | -0.01149 | | | | | | | | | . | | . | | | | | | | | | | | |
| 4 | 0.00598 | | | | | | | | | . | | . | | | | | | | | | | | |
| 5 | -0.00605 | | | | | | | | | . | | . | | | | | | | | | | | |
| 6 | -0.01601 | | | | | | | | | . | | . | | | | | | | | | | | |
| 7 | 0.02135 | | | | | | | | | . | | . | | | | | | | | | | | |
| 8 | 0.06300 | | | | | | | | | . | | * | | | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | | |
|-----|--------|------------------|-------|-------|--------|--------|--------|-------|-------|--|--|--|
| Lag | Square | DF | Prob | | | | | | | | | |
| 6 | 692.35 | 6 | 0.000 | 0.482 | -0.167 | -0.379 | -0.192 | 0.055 | 0.131 | | | |

Y5

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 |
|-----|---------|---------|----|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|
| 0 | 1.77591 | 1.00000 | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 0.88862 | 0.50037 | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | -0.0056 | -.00314 | | | | | | | | | | . | | . | | | | | | | | | |
| 3 | -0.074 | -.04169 | | | | | | | | | | * | | . | | | | | | | | | |
| 4 | -0.0503 | -.02831 | | | | | | | | | | * | | . | | | | | | | | | |
| 5 | 0.03023 | 0.01702 | | | | | | | | | | . | | . | | | | | | | | | |
| 6 | 0.0327 | 0.01841 | | | | | | | | | | . | | . | | | | | | | | | |
| 7 | 0.00366 | 0.00206 | | | | | | | | | | . | | . | | | | | | | | | |
| 8 | 0.06513 | 0.03667 | | | | | | | | | | . | | * | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|------|---|-------|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.77909 | | | | | | | | | | | | ***** | . | | | | | | | | | |
| 2 | 0.58550 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 3 | -0.42620 | | | | | | | | | | * | | ***** | . | | | | | | | | | |
| 4 | 0.30980 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 5 | -0.20537 | | | | | | | | | | **** | | . | | | | | | | | | | |
| 6 | 0.10971 | | | | | | | | | | . | | ** | | | | | | | | | | |
| 7 | -0.04141 | | | | | | | | | | * | | . | | | | | | | | | | |
| 8 | 0.00364 | | | | | | | | | | . | | . | | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|-------|---|-------|---|---|---|---|---|---|---|---|---|--|
| 1 | 0.50037 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 2 | -0.33819 | | | | | | | | | | ***** | | . | | | | | | | | | | |
| 3 | 0.19526 | | | | | | | | | | . | | **** | | | | | | | | | | |
| 4 | -0.15162 | | | | | | | | | | *** | | . | | | | | | | | | | |
| 5 | 0.14907 | | | | | | | | | | . | | *** | | | | | | | | | | |
| 6 | -0.11616 | | | | | | | | | | ** | | . | | | | | | | | | | |
| 7 | 0.09129 | | | | | | | | | | . | | ** | | | | | | | | | | |
| 8 | -0.00938 | | | | | | | | | | . | | . | | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | | |
|-----|--------|------------------|-------|-------|--------|--------|--------|-------|-------|--|--|--|
| Lag | Square | DF | Prob | | | | | | | | | |
| 6 | 381.10 | 6 | 0.000 | 0.500 | -0.003 | -0.042 | -0.028 | 0.017 | 0.018 | | | |

Y6

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|-------|----|---|---|-------|---|---|---|---|---|---|---|---|---|--|
| 0 | 1.3101 | 1.00000 | | | | | | | | | | | | ***** | | | | | | | | | | |
| 1 | -0.1529 | -.11669 | | | | | | | | | ** | | . | | | | | | | | | | | |
| 2 | -0.4136 | -.31571 | | | | | | | | ***** | | . | | | | | | | | | | | | |
| 3 | -0.0558 | -.04262 | | | | | | | | * | | . | | | | | | | | | | | | |
| 4 | -0.0349 | -.02664 | | | | | | | | * | | . | | | | | | | | | | | | |
| 5 | 0.0462 | 0.03526 | | | | | | | | . | | * | | | | | | | | | | | | |
| 6 | 0.02675 | 0.02042 | | | | | | | | . | | . | | | | | | | | | | | | |
| 7 | -0.0657 | -.05012 | | | | | | | | * | | . | | | | | | | | | | | | |
| 8 | 0.0442 | 0.03374 | | | | | | | | . | | * | | | | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|----|----|-----|-------|-------|-------|-------|---|---|---|---|---|---|---|---|---|--|
| 1 | 0.46020 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 2 | 0.52259 | | | | | | | | | . | | ***** | | | | | | | | | | | |
| 3 | 0.33470 | | | | | | | | . | | ***** | | | | | | | | | | | | |
| 4 | 0.28528 | | | | | | | . | | ***** | | | | | | | | | | | | | |
| 5 | 0.17483 | | | | | | . | | *** | | | | | | | | | | | | | | |
| 6 | 0.11407 | | | | | . | | ** | | | | | | | | | | | | | | | |
| 7 | 0.08309 | | | | . | | ** | | | | | | | | | | | | | | | | |
| 8 | 0.01913 | | | . | | . | | | | | | | | | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|----|----|----|------|-------|----|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.11669 | | | | | | | | | ** | | . | | | | | | | | | | | |
| 2 | -0.33387 | | | | | | | | ***** | | . | | | | | | | | | | | | |
| 3 | -0.14914 | | | | | | | | *** | | . | | | | | | | | | | | | |
| 4 | -0.19317 | | | | | | | **** | | . | | | | | | | | | | | | | |
| 5 | -0.08641 | | | | | | ** | | . | | | | | | | | | | | | | | |
| 6 | -0.08499 | | | | | ** | | . | | | | | | | | | | | | | | | |
| 7 | -0.11042 | | | | ** | | . | | | | | | | | | | | | | | | | |
| 8 | -0.02902 | | | * | | . | | | | | | | | | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | |
|-----|--------|------------------|-------|--------|--------|--------|--------|-------|-------|
| Lag | Square | DF | Prob | | | | | | |
| 6 | 176.67 | 6 | 0.000 | -0.117 | -0.316 | -0.043 | -0.027 | 0.035 | 0.020 |

Y7

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|--|
| 0 | 1.05471 | 1.00000 | | | | | | | | | | | | ***** | | | | | | | | | | |
| 1 | 0.02858 | 0.02710 | | | | | | | | | | . | * | | | | | | | | | | | |
| 2 | -0.0025 | -.00234 | | | | | | | | | | . | . | | | | | | | | | | | |
| 3 | -0.0332 | -.03150 | | | | | | | | | | * | . | | | | | | | | | | | |
| 4 | -0.0234 | -.02215 | | | | | | | | | | . | . | | | | | | | | | | | |
| 5 | 0.01823 | 0.01729 | | | | | | | | | | . | . | | | | | | | | | | | |
| 6 | 0.02353 | 0.02231 | | | | | | | | | | . | . | | | | | | | | | | | |
| 7 | -0.0265 | -.02510 | | | | | | | | | | * | . | | | | | | | | | | | |
| 8 | 0.03498 | 0.03316 | | | | | | | | | | . | * | | | | | | | | | | | |

"," marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | -0.02879 | | | | | | | | | | * | . | | | | | | | | | | | |
| 2 | -0.00014 | | | | | | | | | | . | . | | | | | | | | | | | |
| 3 | 0.03119 | | | | | | | | | | . | * | | | | | | | | | | | |
| 4 | 0.02154 | | | | | | | | | | . | . | | | | | | | | | | | |
| 5 | -0.01921 | | | | | | | | | | . | . | | | | | | | | | | | |
| 6 | -0.02185 | | | | | | | | | | . | . | | | | | | | | | | | |
| 7 | 0.02939 | | | | | | | | | | . | * | | | | | | | | | | | |
| 8 | -0.03521 | | | | | | | | | | * | . | | | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 1 | 0.02710 | | | | | | | | | | . | * | | | | | | | | | | | |
| 2 | -0.00307 | | | | | | | | | | . | . | | | | | | | | | | | |
| 3 | -0.03137 | | | | | | | | | | * | . | | | | | | | | | | | |
| 4 | -0.02049 | | | | | | | | | | . | . | | | | | | | | | | | |
| 5 | 0.01832 | | | | | | | | | | . | . | | | | | | | | | | | |
| 6 | 0.02035 | | | | | | | | | | . | . | | | | | | | | | | | |
| 7 | -0.02759 | | | | | | | | | | * | . | | | | | | | | | | | |
| 8 | 0.03539 | | | | | | | | | | . | * | | | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | |
|-----|--------|------------------|-------|-------|--------|--------|--------|-------|-------|
| Lag | Square | DF | Prob | | | | | | |
| 6 | 4.55 | 6 | 0.603 | 0.027 | -0.002 | -0.031 | -0.022 | 0.017 | 0.022 |

Y8

Autocorrelations

| Lag | Covar | Corr | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|---------|---------|----|---|---|---|---|---|---|---|---|---|---|-------|-------|---|---|---|---|---|---|---|---|--|
| 0 | 2.75902 | 1.00000 | | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 2.08823 | 0.75687 | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | 1.21753 | 0.44129 | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 3 | 0.67832 | 0.24585 | | | | | | | | | | . | | **** | | | | | | | | | | |
| 4 | 0.40167 | 0.14558 | | | | | | | | | | . | | *** | | | | | | | | | | |
| 5 | 0.28744 | 0.10418 | | | | | | | | | | . | | ** | | | | | | | | | | |
| 6 | 0.21441 | 0.07771 | | | | | | | | | | . | | ** | | | | | | | | | | |
| 7 | 0.15825 | 0.05736 | | | | | | | | | | . | | * | | | | | | | | | | |
| 8 | 0.15586 | 0.05649 | | | | | | | | | | . | | * | | | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | | |
|-----|-------------|----|---|---|---|---|---|---|---|---|----|---|---|-------|-------|---|---|---|---|---|---|---|--|--|
| 1 | -0.67877 | | | | | | | | | | | | | | ***** | | | | | | | | | |
| 2 | 0.26582 | | | | | | | | | | | . | | ***** | | | | | | | | | | |
| 3 | -0.09853 | | | | | | | | | | ** | | . | | | | | | | | | | | |
| 4 | 0.05250 | | | | | | | | | | . | | * | | | | | | | | | | | |
| 5 | -0.02564 | | | | | | | | | | * | | . | | | | | | | | | | | |
| 6 | -0.01058 | | | | | | | | | | . | | . | | | | | | | | | | | |
| 7 | 0.02696 | | | | | | | | | | . | | * | | | | | | | | | | | |
| 8 | -0.01597 | | | | | | | | | | . | | . | | | | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | | |
|-----|-------------|----|---|---|---|---|---|---|---|---|---|---|----|---|-------|---|---|---|---|---|---|---|--|--|
| 1 | 0.75687 | | | | | | | | | | | | . | | ***** | | | | | | | | | |
| 2 | -0.30802 | | | | | | | | | | | | | | ***** | | | | | | | | | |
| 3 | 0.10891 | | | | | | | | | | . | | ** | | | | | | | | | | | |
| 4 | -0.01306 | | | | | | | | | | . | | . | | | | | | | | | | | |
| 5 | 0.03785 | | | | | | | | | | . | | * | | | | | | | | | | | |
| 6 | -0.01800 | | | | | | | | | | . | | . | | | | | | | | | | | |
| 7 | 0.01313 | | | | | | | | | | . | | . | | | | | | | | | | | |
| 8 | 0.03612 | | | | | | | | | | . | | * | | | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | | |
|-----|---------|------------------|-------|-------|-------|-------|-------|-------|-------|--|--|--|
| Lag | Square | DF | Prob | | | | | | | | | |
| 6 | 1302.24 | 6 | 0.000 | 0.757 | 0.441 | 0.246 | 0.146 | 0.104 | 0.078 | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 |
|-----|-------------|----|---|---|---|---|---|---|---|---|----|-------|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.43521 | | | | | | | | | . | | ***** | | | | | | | | | | |
| 2 | 0.31370 | | | | | | | | | . | | ***** | | | | | | | | | | |
| 3 | -0.08479 | | | | | | | | | . | ** | | . | | | | | | | | | |
| 4 | -0.05397 | | | | | | | | | . | * | | . | | | | | | | | | |
| 5 | 0.11732 | | | | | | | | | . | | ** | . | | | | | | | | | |
| 6 | 0.00206 | | | | | | | | | . | | | . | | | | | | | | | |
| 7 | 0.02711 | | | | | | | | | . | | * | . | | | | | | | | | |
| 8 | 0.09302 | | | | | | | | | . | | ** | . | | | | | | | | | |

Autocorrelation Check for White Noise

| To | Chi | Autocorrelations | | | | | | | | | |
|-----|--------|------------------|-------|-------|-------|-------|-------|-------|-------|--|--|
| Lag | Square | DF | Prob | | | | | | | | |
| 6 | 37.03 | 6 | 0.000 | 0.435 | 0.444 | 0.207 | 0.142 | 0.151 | 0.090 | | |

Looks like AR(2) or MA(2) may fit well.

How to fit an MA ?

Data: 10 12 13 11 9 10 8 9 8 (Mean = 10)

Sum of squares for $\theta = -.5$ $Y_t = e_t - \theta e_{t-1}$

| | | | | | | | | | |
|--------------------------|----|----|----|----|----|------|-------|--------|---------|
| Y_t | 10 | 12 | 13 | 11 | 9 | 10 | 8 | 9 | 8 |
| $y_t = Y_t - 10$ | 0 | 2 | 3 | 1 | -1 | 0 | -2 | -1 | -2 |
| $\hat{y}_t = .5 e_{t-1}$ | 0 | 0 | 1 | 1 | 0 | -0.5 | 0.25 | -1.125 | 0.0625 |
| $e_t = y_t - \hat{y}_t$ | 0 | 2 | 2 | 0 | -1 | 0.5 | -2.25 | 0.125 | -2.0625 |

sum of squared errors = $0^2 + 2^2 + \dots + 2.0625^2 = 18.582$.

| | | | | | | | | | |
|----------|------|----|------|------|------|-------|--------------|------|------|
| θ | .1 | 0 | -.1 | -.2 | -.3 | -.4 | -.5 | -.6 | -.7 |
| SS(err) | 26.7 | 24 | 21.9 | 20.3 | 19.2 | 18.64 | 18.58 | 19.2 | 20.6 |

so $\hat{\theta} \approx -.5$.

A better idea: Minimize $-2 \ln(\text{likelihood}) = -2 \ln(\mathcal{L})$. Works for any ARMA(p,q).

Likelihood based statistics

- $\ln(\mathcal{L}) = \log$ of likelihood: big is good.
- $AIC = -2 \ln(\mathcal{L}) + 2(p+q)$ small is good
Akaike's Information Criterion
- $SBC = -2 \ln(\mathcal{L}) + (p+q) \ln(n)$
Schwartz's Bayesian Criterion
- Compare different models for the same data
Use same estimation method
Do not compare to each other
SBC more conservative.

Example: Baseball winning % for NL West pennant winner

```

Data Pennant;
  retain year 1900;   year=year+1; input winpct @@;
  title h=3 "Baseball";   split = (year>1968);   if year > 1920;
cards;
647 741 650 693 686 763 704 643 724 675 647 682 664 614 592 610
636 651 686 604 614 604 621 608 621 578 610 617 645 597 656 584
599 621 649 597 625 586 630 654 649 688 682 682 636 628 610 595
630 591 624 627 682 630 641 604 617 597 564 617 604 624 611 574
599 586 627 599 574 630 556 617 611 630 667 630 605 586 559 571
632 549 562 568 586 593 556 584 568 562 580 605 642
;
title2 "West";
proc arima; i var=winpct;
  e p=1 ml itprint grid; e p=2 ml; e q=2 ml;
  e q=1 p=2 ml; run;
    
```

Box-Ljung on original data (identify statement)

| Baseball West The ARIMA Procedure Autocorrelation Check for White Noise | | | | | | | | | |
|----------------------------------------------------------------------------------|--------|------|----------------------------|-------|-------|--------|--------|--------|-------|
| To | Chi- | Pr > | -----Autocorrelations----- | | | | | | |
| Lag | Square | DF | ChiSq | | | | | | |
| 6 | 37.03 | 6 | <.0001 | 0.435 | 0.444 | 0.207 | 0.142 | 0.151 | 0.090 |
| 12 | 46.99 | 12 | <.0001 | 0.122 | 0.141 | 0.131 | 0.130 | 0.151 | 0.153 |
| 18 | 51.31 | 18 | <.0001 | 0.097 | 0.049 | -0.079 | -0.087 | -0.103 | 0.094 |

Estimate statement with itprint, grid option

| Preliminary Estimation | | | | | | |
|-----------------------------------------|---------------------------------------|----------|---------|----------|---------|----------|
| Initial Autoregressive Estimates | | | | | | |
| | Estimate | | | | | |
| 1 | 0.43521 | | | | | |
| Constant Term Estimate | | 344.7338 | | | | |
| White Noise Variance Est | | 831.0392 | | | | |
| Conditional Least Squares Estimation | | | | | | |
| Iteration | SSE | MU | AR1,1 | Constant | Lambda | R Crit |
| 0 | 60476 | 610.3699 | 0.43521 | 344.7338 | 0.00001 | 1 |
| 1 | 60470 | 610.7646 | 0.44120 | 341.297 | 1E-6 | 0.010208 |
| 2 | 60470 | 610.7702 | 0.44123 | 341.281 | 1E-7 | 0.000114 |
| Maximum Likelihood Estimation | | | | | | |
| Iter | Loglike | MU | AR1,1 | Constant | Lambda | R Crit |
| 0 | -348.94925 | 610.7702 | 0.44123 | 341.281 | 0.00001 | 1 |
| 1 | -348.94762 | 610.7347 | 0.43512 | 344.9901 | 1E-6 | 0.006768 |
| 2 | -348.94762 | 610.7344 | 0.43524 | 344.9156 | 1E-7 | 0.000135 |
| ARIMA Estimation Optimization Summary | | | | | | |
| Estimation Method | Maximum Likelihood | | | | | |
| Parameters Estimated | 2 | | | | | |
| Termination Criteria | Maximum Relative Change in Estimates | | | | | |
| Iteration Stopping Value | 0.001 | | | | | |
| Criteria Value | 0.000279 | | | | | |
| Alternate Criteria | Relative Change in Objective Function | | | | | |
| Alternate Criteria Value | 3.086E-9 | | | | | |
| Maximum Absolute Value of Gradient | 9.014848 | | | | | |
| R-Square Change from Last Iteration | 0.000135 | | | | | |
| Objective Function | Log Gaussian Likelihood | | | | | |
| Objective Function Value | -348.948 | | | | | |
| Marquardt's Lambda Coefficient | 1E-7 | | | | | |
| Numerical Derivative Perturbation Delta | 0.001 | | | | | |
| Iterations | 2 | | | | | |

| Maximum Likelihood Estimation | | | | | | | | | | |
|-------------------------------------|------------|----------------|------------|----------------------------|-------|--------|--------|--------|--------|--|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag | | | | | |
| MU | 610.73440 | 5.97709 | 102.18 | <.0001 | 0 | | | | | |
| AR1,1 | 0.43524 | 0.10725 | 4.06 | <.0001 | 1 | | | | | |
| Constant Estimate | | 344.9156 | | | | | | | | |
| Variance Estimate | | 851.6998 | | | | | | | | |
| Std Error Estimate | | 29.1839 | | | | | | | | |
| AIC | | 701.8952 | | | | | | | | |
| SBC | | 706.4762 | | | | | | | | |
| Number of Residuals | | 73 | | | | | | | | |
| Correlations of Parameter Estimates | | | | | | | | | | |
| Parameter | MU | AR1,1 | | | | | | | | |
| MU | 1.000 | 0.024 | | | | | | | | |
| AR1,1 | 0.024 | 1.000 | | | | | | | | |
| Autocorrelation Check of Residuals | | | | | | | | | | |
| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | | |
| 6 | 9.32 | 5 | 0.0969 | -0.129 | 0.309 | -0.012 | 0.009 | 0.090 | -0.015 | |
| 12 | 11.03 | 11 | 0.4405 | 0.044 | 0.069 | 0.045 | 0.033 | 0.066 | 0.074 | |
| 18 | 17.65 | 17 | 0.4111 | 0.043 | 0.057 | -0.106 | -0.037 | -0.159 | 0.159 | |
| 24 | 20.25 | 23 | 0.6266 | 0.017 | 0.007 | 0.114 | -0.071 | 0.074 | 0.023 | |
| Log-likelihood Surface on Grid | | | | | | | | | | |
| Near Estimates: AR1,1 (winpct) | | | | | | | | | | |
| MU (winpct) | 0.43024 | 0.43524 | 0.44024 | | | | | | | |
| | 610.72940 | -348.95 | -348.95 | -348.95 | | | | | | |
| | 610.73440 | -348.95 | -348.95 | -348.95 | | | | | | |
| | 610.73940 | -348.95 | -348.95 | -348.95 | | | | | | |
| Model for variable winpct | | | | | | | | | | |
| Estimated Mean | | 610.7344 | | | | | | | | |
| Autoregressive Factors | | | | | | | | | | |
| Factor 1: 1 - 0.43524 B**(1) | | | | | | | | | | |

B= "backshift" $B(Y_t)=Y_{t-1}$.

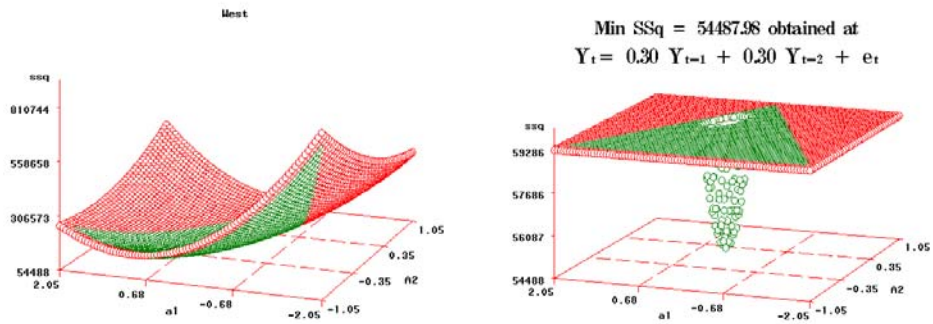
$$(1-0.4352B)(Y_t-\mu)=e_t \Rightarrow (Y_t-\mu)-0.4352(Y_{t-1}-\mu)=e_t \Rightarrow (Y_t-\mu)=0.4352(Y_{t-1}-\mu) + e_t$$

| Maximum Likelihood Estimation | | | | | | | | | |
|----------------------------------------------|------------|----------------|------------|----------------------------|-------|--------|--------|--------|--------|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag | | | | |
| MU | 610.94760 | 8.00314 | 76.34 | <.0001 | 0 | | | | |
| AR1,1 | 0.29680 | 0.11428 | 2.60 | 0.0094 | 1 | | | | |
| AR1,2 | 0.30822 | 0.11431 | 2.70 | 0.0070 | 2 | | | | |
| Constant Estimate | | 241.3102 | | | | | | | |
| Variance Estimate | | 779.1167 | | | | | | | |
| Std Error Estimate | | 27.91266 | | | | | | | |
| AIC | | 696.5505 | | | | | | | |
| SBC | | 703.4218 | | | | | | | |
| Number of Residuals | | 73 | | | | | | | |
| Correlations of Parameter Estimates | | | | | | | | | |
| Parameter | MU | AR1,1 | AR1,2 | | | | | | |
| MU | 1.000 | 0.017 | 0.016 | | | | | | |
| AR1,1 | 0.017 | 1.000 | -0.443 | | | | | | |
| AR1,2 | 0.016 | -0.443 | 1.000 | | | | | | |
| Autocorrelation Check of Residuals | | | | | | | | | |
| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
| 6 | 1.99 | 4 | 0.7379 | 0.033 | 0.067 | -0.074 | -0.101 | 0.034 | -0.054 |
| 12 | 3.35 | 10 | 0.9720 | 0.007 | 0.057 | 0.021 | 0.006 | 0.063 | 0.087 |
| 18 | 13.71 | 16 | 0.6200 | 0.067 | 0.052 | -0.125 | -0.163 | -0.171 | 0.168 |
| 24 | 16.38 | 22 | 0.7965 | 0.061 | 0.022 | 0.100 | -0.057 | 0.088 | 0.008 |
| Model for variable winpct | | | | | | | | | |
| Estimated Mean | | 610.9476 | | | | | | | |
| Autoregressive Factors | | | | | | | | | |
| Factor 1: 1 - 0.2968 B**(1) - 0.30822 B**(2) | | | | | | | | | |

$m^2 - 0.2968m - 0.30822 = (m - 0.7231)(m + 4.263)$. Fitted model is stationary !!

$$Y_t = \frac{1}{(1 - 0.7231B)(1 + 4.263B)} e_t = \left[\frac{A}{(1 - 0.7231B)} + \frac{1-A}{(1 + 4.263B)} \right] e_t$$

$$A = .7231 / (.7231 + 4.263)$$



| Maximum Likelihood Estimation | | | | | | | | | |
|-----------------------------------------------|------------|----------------|------------|----------------------------|--------|--------|--------|--------|--------|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag | | | | |
| MU | 610.79315 | 5.63555 | 108.38 | <.0001 | 0 | | | | |
| MA1,1 | -0.30846 | 0.10851 | -2.84 | 0.0045 | 1 | | | | |
| MA1,2 | -0.43694 | 0.10949 | -3.99 | <.0001 | 2 | | | | |
| Constant Estimate | | 610.7931 | | | | | | | |
| Variance Estimate | | 779.7421 | | | | | | | |
| Std Error Estimate | | 27.92386 | | | | | | | |
| AIC | | 696.6768 | | | | | | | |
| SBC | | 703.5482 | | | | | | | |
| Number of Residuals | | 73 | | | | | | | |
| Autocorrelation Check of Residuals | | | | | | | | | |
| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
| 6 | 2.46 | 4 | 0.6517 | 0.046 | 0.059 | 0.139 | 0.044 | 0.067 | -0.004 |
| 12 | 4.58 | 10 | 0.9172 | 0.041 | 0.098 | 0.046 | 0.024 | 0.083 | 0.062 |
| 18 | 12.38 | 16 | 0.7177 | 0.038 | 0.077 | -0.055 | -0.139 | -0.125 | 0.186 |
| 24 | 14.23 | 22 | 0.8932 | 0.028 | -0.013 | 0.102 | -0.045 | 0.063 | 0.007 |
| Model for variable winpct | | | | | | | | | |
| Estimated Mean | | 610.7931 | | | | | | | |
| Moving Average Factors | | | | | | | | | |
| Factor 1: 1 + 0.30846 B**(1) + 0.43694 B**(2) | | | | | | | | | |

$$m^2 + 0.30846m + 0.43694 = (m - A)(m - A^*)$$

$$m^2 + 0.30846m + 0.43694 = (m - A)(m - A^*)$$

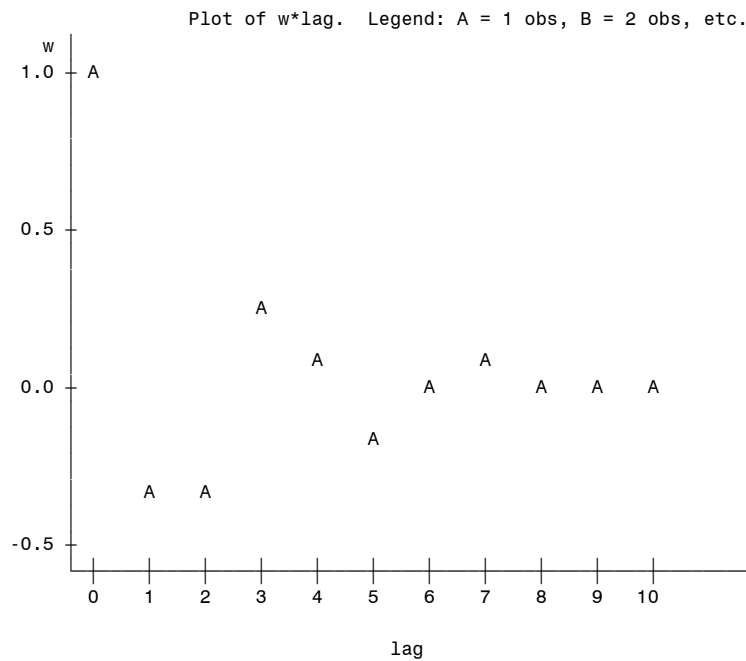
A, A* complex pair with $|A| = \sqrt{0.43694} = 0.6610 < 1$. Fitted model is invertible!!

$$\begin{aligned} e_t &= w_0(Y_{t-\mu}) + w_1(Y_{t-1-\mu}) + w_2(Y_{t-2-\mu}) + w_3(Y_{t-3-\mu}) + w_4(Y_{t-4-\mu}) + \dots \\ .30846 e_{t-1} &= .3046[w_0(Y_{t-1-\mu}) + w_1(Y_{t-2-\mu}) + w_2(Y_{t-3-\mu}) + w_3(Y_{t-4-\mu}) + w_4(Y_{t-5-\mu}) + \dots \\ .43694 e_{t-2} &= .43694[w_0(Y_{t-2-\mu}) + w_1(Y_{t-3-\mu}) + w_2(Y_{t-4-\mu}) + w_3(Y_{t-5-\mu}) + \dots \\ (\text{sum}) &= 1(Y_{t-\mu}) \end{aligned}$$

so we must have (with $w_{-1}=0$) $w_0=1$ and $w_j + .30846w_{j-1} + .43694w_{j-2} = 0$

Generate the ws: observe damped cyclical behavior - typical of complex roots!!

```
data weights;
w1 = 1; w2=0;w=1;lag=0; output;
do lag = 1 to 10;
  w = -.30846*w1 - .43694*w2;
output; retain; w2=w1;w1=w; end;
proc plot; plot w*lag/hpos=60 vpos=20;
title "weights"; run;
```



Maximum Likelihood Estimation

| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag |
|---------------------|-----------|----------------|----------|----------------|-----|
| MU | 610.94413 | 7.77268 | 78.60 | <.0001 | 0 |
| MA1,1 | -0.17953 | 0.37373 | -0.48 | 0.6310 | 1 |
| AR1,1 | 0.13527 | 0.34434 | 0.39 | 0.6944 | 1 |
| AR1,2 | 0.38109 | 0.16772 | 2.27 | 0.0231 | 2 |
| Constant Estimate | | | 295.4738 | | |
| Variance Estimate | | | 785.8795 | | |
| Std Error Estimate | | | 28.03354 | | |
| AIC | | | 698.1473 | | |
| SBC | | | 707.3092 | | |
| Number of Residuals | | | 73 | | |

Autocorrelation Check of Residuals

| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
|--------|------------|----|------------|----------------------------|-------|--------|--------|--------|--------|
| 6 | 1.64 | 3 | 0.6494 | 0.012 | 0.052 | -0.030 | -0.104 | 0.059 | -0.052 |
| 12 | 2.92 | 9 | 0.9673 | 0.008 | 0.064 | 0.022 | 0.003 | 0.063 | 0.077 |
| 18 | 12.95 | 15 | 0.6058 | 0.065 | 0.053 | -0.112 | -0.156 | -0.168 | 0.177 |
| 24 | 15.88 | 21 | 0.7763 | 0.054 | 0.012 | 0.105 | -0.069 | 0.093 | 0.004 |

Model for variable winpct

Estimated Mean 610.9441

Autoregressive Factors

Factor 1: 1 - 0.13527 B**(1) - 0.38109 B**(2)

Moving Average Factors

Factor 1: 1 + 0.17953 B**(1)

| | AIC | SBC |
|---------------------|--------------|--------------|
| AR(1) | 701.8952 | 706.4762 |
| AR(2) | 696.5505 *** | 703.4218 *** |
| MA(2) | 696.6768 | 703.5482 |
| ARMA(2,1) | 698.1473 | 707.3092 |
| ARMA(1,1) not shown | 699.4642 | 706.3356 |

Models for nonstationary data

(some excerpts from SAS for forecasting Time Series)

Example: AR(2) model

$$Y_{t-\mu} = \alpha_1(Y_{t-1-\mu}) + \alpha_2(Y_{t-2-\mu}) + e_t$$

same as model in differences and ONE lag level ($Y_{t-1-\mu}$).

$$Y_t - Y_{t-1} = - (1-\alpha_1 - \alpha_2)(Y_{t-1-\mu}) - \alpha_2(Y_{t-1} - Y_{t-2}) + e_t$$

Characteristic equation

$$1 - \alpha_1 M - \alpha_2 M^2 = 0$$

$M=1$ a root implies $(1-\alpha_1 - \alpha_2)=0$ so $(1-\alpha_1 - \alpha_2)(Y_{t-1-\mu}) = 0$ (no mean in equation now)

Forecasts no longer revert to the mean. I prefer descending powers

$$m^2 - \alpha_1 m - \alpha_2 = 0$$

but again, if $m=1$ is a root then $(1-\alpha_1 - \alpha_2)=0$ which is all that matters at the moment. In general $m=1/M$ relates the nonzero roots in the two representations.

Estimation:

OLS (ordinary least squares)

Regression of $Y_t - Y_{t-1}$ on Y_{t-1} and $(Y_{t-1} - Y_{t-2})$ with an intercept.

Test based on coefficient or t test on the Y_{t-1} term as a test of the null hypothesis that the series has a unit root nonstationarity.

If all roots M exceed 1 in magnitude ($|m| < 1$) the coefficient of $(Y_{t-1-\mu})$ will be negative, suggesting a one tailed test to the left if stationarity is the alternative.

Major problem: neither the estimated coefficient of $(Y_{t-1-\mu})$ nor its t test have standard distributions, even when the sample size becomes very large. (Can still test but need a new distribution for the test statistics).

Dickey and Fuller (1979, 1981) studied 3 models. The leftmost column of the following table shows the regressions they studied. Rightmost is model. Here $Y_t - Y_{t-1} = \nabla Y_t$ $(1-B)Y_t$ denotes a first difference.

| | |
|------------------------------------------------------|----------------------------------------------------------------|
| regress ∇Y_t on these: | AR(1) in deviations form |
| $Y_{t-1}, \nabla Y_{t-1} \dots \nabla Y_{t-k}$ | $Y_t = \rho Y_{t-1} + e_t$ |
| $Y_{t-1}, 1, \nabla Y_{t-1} \dots \nabla Y_{t-k}$ | $Y_{t-\mu} = \rho(Y_{t-1-\mu}) + e_t$ |
| $Y_{t-1}, 1, t, \nabla Y_{t-1} \dots \nabla Y_{t-k}$ | $Y_{t-\alpha-\beta t} = \rho(Y_{t-1-\alpha-\beta(t-1)}) + e_t$ |

| | |
|-------------------------------------------------------------------------|----------------------------|
| AR(1) in regression form | $H_0: \rho=1$ |
| $\nabla Y_t = (\rho-1)Y_{t-1} + e_t$ | $\nabla Y_t = e_t$ |
| $\nabla Y_t = (1-\rho)\mu + (\rho-1)Y_{t-1} + e_t$ | $\nabla Y_t = e_t$ |
| $\nabla Y_t = (1-\rho)(\alpha+\beta t) + \beta + (\rho-1)Y_{t-1} + e_t$ | $\nabla Y_t = \beta + e_t$ |

The lagged differences -> "augmenting lags," tests -> "Augmented Dickey-Fuller" or "ADF" tests. Deviations from μ form shows that if $|\rho| < 1$ and we have appropriate starting values, the expected value of Y_t is 0, μ , or $\alpha + \beta t$ depending on which model is assumed.

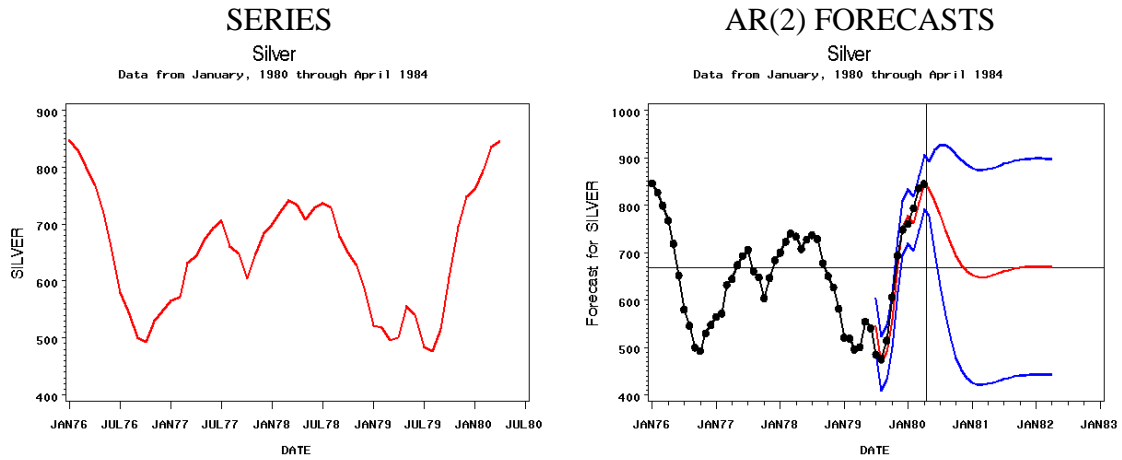
Fit the first model only if you know the mean of your data is 0 (for example, Y_t might already be a difference of some observed variable). Use the third model if you suspect a regular trend up or down in your data. If you fit the third model when β is really 0, your tests will be valid, but not as powerful as those from the second model.

The parameter β represents a trend slope when $|\rho| < 1$ and is called a "drift" when $\rho = 1$. Note that for known parameters and n data points, the forecast of Y_{n+L} would be $\alpha + \beta(n+L) + \rho^L(Y_n - \alpha - \beta n)$ for $|\rho| < 1$ with forecast error variance $(1 + \rho^2 + \dots + \rho^{2L-2})\sigma^2$. As L increases, $(1 + \rho^2 + \dots + \rho^{2L-2})\sigma^2$ approaches $\sigma^2/(1 - \rho^2)$, the variance of Y around the trend. However, if $\rho = 1$ the L step ahead forecast is $Y_n + \beta L$ with forecast error variance $L\sigma^2$, so that the error variance increases without bound in this case. In both cases, the forecasts have a component that increases at the linear rate β .

Features

- (1) Distributions for the coefficients of Y_{t-1} , 1, and t are all nonstandard.
Critical values and theory: Fuller (1996).
- (2) Coefficients of the lagged differences $\nabla Y_{t,j}$ have limiting normal distributions.
Can use F to test backwards on lags. Alastair Hall (1994) shows this is good method.
- (3) Coefficients of Y_{t-1} and the associated t tests have distributions that differ among the three regressions and are nonstandard.
- (4) t test statistics have the same limit distributions no matter how many augmenting lags are used.

Example:stocks of silver on the New York Commodities Exchange.



DEL is difference, DELi is ith lag and LSILVER is lagged level of silver.

```
PROC REG; MODEL DEL=LSILVER DEL1 DEL2 DEL3 DEL4;
TEST DEL2=0, DEL3=0, DEL4=0;
PROC REG; MODEL DEL=LSILVER DEL1;
```

Output:

| | | | | | |
|-------------------------|--------------|-----------|--------|----------|--------|
| Dependent Variable: DEL | | | | | |
| Test: | Numerator: | 1152.1971 | DF: 3 | F value: | 1.3221 |
| | Denominator: | 871.5178 | DF: 41 | Prob>F: | 0.2803 |

Test involves only the lagged differences - F justified in large samples. No evidence against leaving out all but the first augmenting lag so ...

Output:

| Parameter Estimates | | | | | |
|---------------------|----|--------------------|----------------|-----------------------|-----------|
| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > T |
| INTERCEP | 1 | 75.580727 | 27.36395089 | 2.762 | 0.0082 |
| LSILVER | 1 | -0.117034 | 0.04216383 | -2.776 | 0.0079 |
| DEL1 | 1 | 0.671152 | 0.10806009 | 6.211 | 0.0001 |

P-value 0.0079 < 0.05. Reject unit root? (conclude stationary?) **NO**

PROC REG uses t distribution for p-values! Appropriate 5% left tail critical value -2.86 (Fuller, 1996, pg. 642). Nonstationarity (unit roots) cannot be rejected. PROC ARIMA gives SAME test statistics, CORRECT p-values:

```
PROC ARIMA;
  i var = silver stationarity=(ADF=(1))
```

Output

| Augmented Dickey-Fuller Unit Root Tests | | | | | | | |
|-----------------------------------------|------|----------|----------|-------|----------|------|--------|
| Type | Lags | Rho | Pr < Rho | Tau | Pr < Tau | F | Pr > F |
| Zero Mean | 1 | -0.2461 | 0.6232 | -0.28 | 0.5800 | | |
| Single Mean | 1 | -17.7945 | 0.0121 | -2.78 | 0.0689 | 3.86 | 0.1197 |
| Trend | 1 | -15.1102 | 0.1383 | -2.63 | 0.2697 | 4.29 | 0.3484 |

Every observed data point exceeds 400. Test that assumes a 0 mean (Tau = -0.28) can be ignored! Monte Carlo study (Dickey Proceedings of B&E section ASA) shows that assuming too simple a trend model among the three results in almost no power (almost always fails to reject unit roots). Tests with no lagged (augmenting) differences can also be ignored (REG=> need 1). Tests with too many augmenting lags are valid, but not as powerful (just as throwing unneeded terms into any regression typically reduces power because of the extraneous estimates of 0 (those coefficients). "Normalized bias" means $n(\hat{\rho}-1)$ in an AR(1) model with coefficient ρ . Unusual! usually expect normalization $\sqrt{n}(\hat{\rho}-\rho)$ which is OK if $|\rho|<1$ In that case $\sqrt{n}(\hat{\rho}-\rho)$ has normal distribution for large samples with mean 0 and variance $1-\rho^2$.

AR(2) with roots ρ and m :

$$Y_t - Y_{t-1} = - (1-\alpha_1 - \alpha_2)(Y_{t-1} - \mu) - \alpha_2(Y_{t-1} - Y_{t-2}) + e_t$$

$$Y_t - Y_{t-1} = - (1-m)(1-\rho)(Y_{t-1} - \mu) + m\rho(Y_{t-1} - Y_{t-2}) + e_t$$

Coefficient of Y_{t-1} is $-(1-\rho)(1-m)$ which is 0 if $\rho=1$. Coefficient of $(Y_{t-1} - Y_{t-2})$, 0.671152 in the silver example, is an estimate of m , if $\rho=1$, so it is not surprising that an adjustment using that statistic is required to get a test statistic that behaves like $n(\hat{\rho}-1)$ under $H_0: \rho=1$. Specifically you divide the lag 1 coefficient (-0.117034) by $(1-0.671152)$ then multiply by n . Similar adjustments can be made in higher order processes. For the silver data $50(-0.117034)/(1-0.671152) = -17.7945$ is shown in the printout and has p-value less than 0.05.

Simulated size and power results (Dickey, 1984) => tau tests preferable to normalized bias tests. Tau test above, -2.78, has p-value exceeding 0.05. Fails to provide significant

evidence at the usual 0.05 level against the unit root null hypothesis. The F type statistics are discussed in Dickey and Fuller (1981). If interest lies only in inference about ρ , there is no advantage to using the F statistics, which include restrictions on the intercept and trend as a part of H_0 . F tests joint hypothesis, e.g. intercept is 0 and $\rho=1$

Our 50 observations have no apparent trend. Use model with constant mean. Model with linear trend is be valid and would guard against any unrecognized linear trend but is less powerful if no trend exists. Test with validity and good statistical power requires appropriate decisions about the model, in terms of lags and trends. Of course! - Any statistical hypothesis test requires a realistic model for the data.

Summary and additional points:

- (1) Ignorance unit roots, trends-> inappropriate mean reverting forecasts
- (2) p-values produced from usual (stationarity) distributions quite misleading when unit roots are in fact present (silver & closing price p-values too small if t distribution used)
- (3) Regression of differences on trend terms, lagged level, and lagged differences, the usual (t and F) distributions are appropriate for lagged differences in large samples.
- (4) Seasonal dummy variables in a model do not change the large sample (limit) behavior of the unit root tests discussed here. (Dickey, Bell, and Miller (1986))

The series $Y_t = 2.8Y_{t-1} + 0.6Y_{t-2} + 0.8Y_{t-3} + e_t$ uses the operator $(1 - 2.8B - 0.6B^2 - 0.8B^3) = (1-B)(1-B)(1-.8B)$ and so has 2 unit roots, thus $W_t = Y_t - Y_{t-1}$ has a single unit root. Pantula and Dickey (1987) point this out and suggest that if k is an upper bound on the number of unit roots, then testing the $k-1$ differenced series for an additional unit root, then the $k-2$ differenced series, etc. is a sequential strategy that uses already existing tests and has good size and power properties.

Effect of differencing on forecast error variances:

Let $W_t = Y_t - Y_{t-1}$ for some time series Y_t and suppose W_t is stationary, for example let $W_t = .8W_{t-1} + e_t$. Now $Y_{n+L} = Y_n + W_{n+1} + \dots + W_{n+L}$.

Using $W_{n+j} = .8^j W_n + e_{n+j} + .8e_{n+j-1} + \dots + .8^{j-1}e_{n+1} = \hat{W}_{n+j} + e_{n+j} + v_1 e_{n+j-1} + \dots + v_{j-1}e_{n+1}$ where the v_j are seen to be the coefficients in the moving average (or "Wold") representation of the series, you see that the forecast error $W_{n+1} + \dots + W_{n+L}$ is $(1+.8+.8^2+\dots+.8^{L-1})e_{n+1} + (1+.8+.8^2+\dots+.8^{L-2})e_{n+2} + \dots + e_{n+L}$ or in general $(1+v_1+\dots+v_{L-1})e_{n+1} + (1+v_1+\dots+v_{L-2})e_{n+2} + \dots + e_{n+L}$ from which the prediction intervals can be computed. Because $(1+.8+.8^2+\dots)=1/(1-.8) = 5$, you can see that the forecast error is approximately $5(e_{n+1} + e_{n+2} + \dots + e_{n+L})$, the point being that the forecast error *variance* increases linearly (essentially, after the first few forecasts). In the general ARIMA(p,1,q) case where the differences are stationary, the series $1+v_1+v_2+\dots$ also converges exponentially fast to a finite limit and again linearly increasing forecast error variances are encountered in the levels of the series.

Data appear to have trend. Should we difference? **Not necessarily!**

Example:

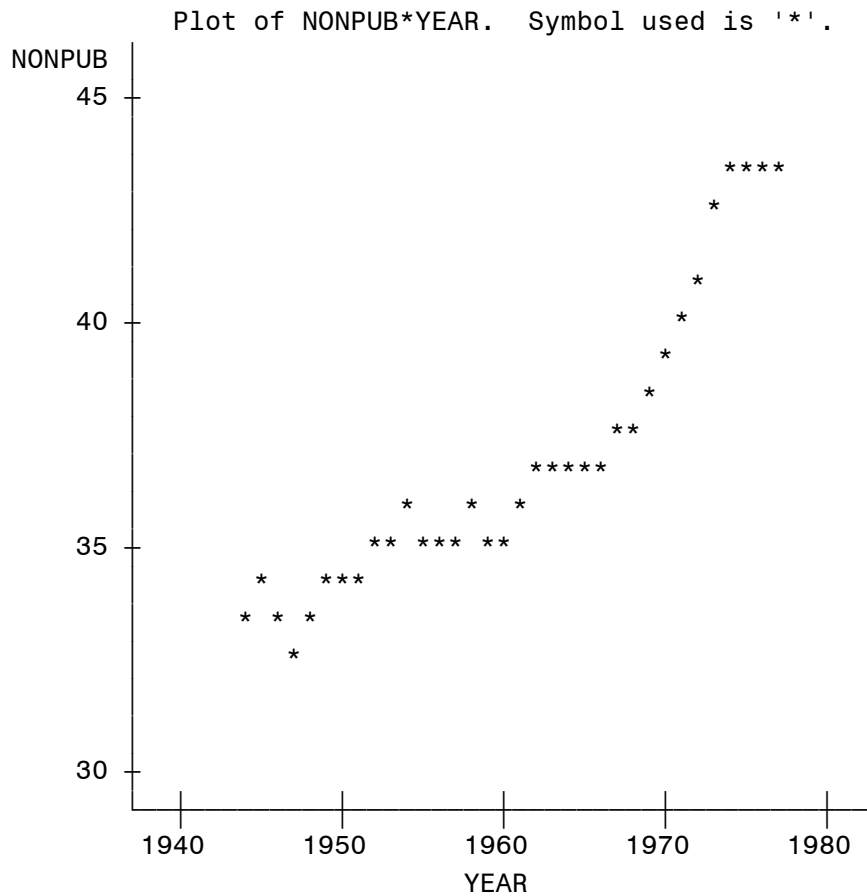
$$Y_t = \alpha + \beta t + e_t \text{ with } e_t \sim N(0, \sigma^2)$$

Best estimate of parameters is least squares regression (basic stat theory).

- Example: Nonproduction workers (%) in publishing industry 1944-1977
(U.S. Bureau of Labor)
- Program:

```

DATA WORKERS;
  INPUT NONPUB @@;
  TITLE 'PUBLISHING AND PRINTING NONPRODUCTION WORKERS';
  TITLE2 '1944-1977';
  RETAIN YEAR 1943;
  YEAR+1;
  CARDS;
33.5 34. 33.5 32.5 33.2 34.1 34 34.3 34.6 35 35.5 35.4 35.1 35.2
35.5 35.3 35.4 35.5 36.6 36.7 36.6 36.4 36.9 37.4 37.7 38.5 39.1
39.7 40.6 42.4 43.2 43.4 43.7 43.5
;
PROC PLOT; PLOT NONPUB*YEAR='*';
PROC ARIMA; IVAR=NONPUB(1) NLAG=5; E Q=1 ML; F LEAD=10; RUN;
    
```



Name of Variable = NONPUB

Period(s) of Differencing 1
 Mean of Working Series 0.30303
 Standard Deviation 0.513741
 Number of Observations 33
 Observation(s) eliminated by differencing 1

Autocorrelations

| Lag | Covariance | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 |
|-----|------------|-------------|----|---|---|---|---|---|---|---|----|---|---|-------|---|---|---|---|---|---|---|---|---|
| 0 | 0.263930 | 1.00000 | | | | | | | | | | | | ***** | | | | | | | | | |
| 1 | 0.082387 | 0.31216 | | | | | | | . | | | | | ***** | . | | | | | | | | |
| 2 | -0.025565 | -.09686 | | | | | | | . | | ** | | | | . | | | | | | | | |
| 3 | 0.0079422 | 0.03009 | | | | | | | . | | | | | * | . | | | | | | | | |
| 4 | 0.034691 | 0.13144 | | | | | | | . | | | | | *** | . | | | | | | | | |
| 5 | 0.010641 | 0.04032 | | | | | | | . | | | | | * | . | | | | | | | | |

"." marks two standard errors

Inverse Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|-------|----|---|---|-------|---|---|---|---|---|---|---|---|--|
| 1 | -0.42926 | | | | | | | | | ***** | | | | | | | | | | | | | |
| 2 | 0.23661 | | | | | | | | . | | | | | ***** | . | | | | | | | | |
| 3 | -0.08486 | | | | | | | | . | | ** | | | | . | | | | | | | | |
| 4 | -0.04814 | | | | | | | | . | | * | | | | . | | | | | | | | |
| 5 | 0.00796 | | | | | | | | . | | | | | | . | | | | | | | | |

Partial Autocorrelations

| Lag | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
|-----|-------------|----|---|---|---|---|---|---|---|---|------|---|---|-------|---|---|---|---|---|---|---|---|--|
| 1 | 0.31216 | | | | | | | | . | | | | | ***** | . | | | | | | | | |
| 2 | -0.21528 | | | | | | | | . | | **** | | | | . | | | | | | | | |
| 3 | 0.15573 | | | | | | | | . | | | | | *** | . | | | | | | | | |
| 4 | 0.05219 | | | | | | | | . | | | | | * | . | | | | | | | | |
| 5 | -0.00997 | | | | | | | | . | | | | | | . | | | | | | | | |

Maximum Likelihood Estimation

| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag |
|-----------|----------|----------------|---------|----------------|-----|
| MU | 0.29996 | 0.12196 | 2.46 | 0.0139 | 0 |
| MA1,1 | -0.45541 | 0.16470 | -2.77 | 0.0057 | 1 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|------------------------------------|------------|----|------------|----------------------------|--------|--------|--------|--------|--------|
| To | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
| Lag 6 | 0.98 | 5 | 0.9644 | -0.025 | -0.091 | 0.024 | 0.115 | -0.028 | -0.038 |
| 12 | 3.65 | 11 | 0.9792 | -0.056 | -0.111 | 0.153 | -0.091 | 0.069 | -0.053 |
| 18 | 7.36 | 17 | 0.9784 | 0.060 | 0.004 | -0.181 | -0.104 | -0.032 | 0.088 |
| 24 | 9.96 | 23 | 0.9915 | 0.005 | -0.020 | 0.068 | -0.005 | -0.121 | 0.057 |

| Model for variable NONPUB | |
|---------------------------|----------|
| Estimated Mean | 0.299963 |
| Period(s) of Differencing | 1 |

| Moving Average Factors | |
|------------------------|--------------------|
| Factor 1: | 1 + 0.45541 B**(1) |

| Forecasts for variable NONPUB | | | | |
|-------------------------------|----------|-----------|-----------------------|---------|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 35 | 43.5664 | 0.4876 | 42.6108 | 44.5221 |
| 36 | 43.8664 | 0.8610 | 42.1788 | 45.5540 |
| 37 | 44.1664 | 1.1158 | 41.9795 | 46.3532 |
| 38 | 44.4663 | 1.3223 | 41.8746 | 47.0580 |
| 39 | 44.7663 | 1.5007 | 41.8249 | 47.7076 |
| 40 | 45.0663 | 1.6600 | 41.8126 | 48.3199 |
| 41 | 45.3662 | 1.8054 | 41.8278 | 48.9047 |
| 42 | 45.6662 | 1.9398 | 41.8642 | 49.4682 |
| 43 | 45.9661 | 2.0656 | 41.9177 | 50.0146 |
| 44 | 46.2661 | 2.1841 | 41.9854 | 50.5468 |

Exponential Smoothing:

If the first differences of a series follow a moving average of order 1, you have

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1} \text{ whence } e_t = (Y_t - Y_{t-1}) + \theta(Y_{t-1} - Y_{t-2}) + \theta^2(Y_{t-2} - Y_{t-3}) + \dots$$

$$\text{or } Y_t = e_t + [(1-\theta) Y_{t-1} + \theta(1-\theta) Y_{t-1} + \theta^2(1-\theta) Y_{t-1} + \dots]$$

$$\text{or } Y_t = e_t + [S_{t-1}] \text{ You see that:}$$

- S_{t-1} is the "best" forecast of Y_t
- S_{t-1} is a weighted average of Y_{t-1}, Y_{t-2}, \dots
- $S_t = (1-\theta)Y_t + \theta S_{t-1}$ = weighted average of Y_t and its forecast
- Replacing Y_j with 0 for $j < 0$ has little effect on S_t once t gets reasonably large.

This is an old fashioned technique called exponential smoothing, attributed to Brown. Originally, the value $1-\theta$, called the "smoothing weight" was picked arbitrarily within the (0,1) interval and an arbitrary value (usually Y_1 itself or the average of the first few Y_s) was used as S_0 . From there $S_t = (1-\theta)Y_t + \theta S_{t-1}$ was used for $t=1$ to n then S_n reported as the forecast of Y_{n+1} (and of all future Y_s) with updating when the next observation came in. The forecast out into the future was a horizontal line. You can feed the smoothed values S_t back into another smoothing loop to produce "double exponentially smoothed" values and the forecasts from these are a line (not horizontal). Seasonal versions by Winters and Holt have also appeared in the literature as has triple exponential smoothing. The later results in quadratic forecasts.

PROC FORECAST in SAS is set up to (as options) do all of these methods. Because all of these are equivalent to certain ARIMA models in which the weights are custom fit to the data, it is difficult to motivate the use of exponential smoothing today. Exponential smoothing is nevertheless still in common use. Quality control is often done in this way, i.e. by keeping track of an exponentially smoothed version of, say, viscosity then sounding a warning when it exceeds some predetermined bounds. Note that an ARIMA(0,1,1) is actually a little more general in that it allows the smoothing weight $1-\theta$ to exceed 1 and a mean (drift) term can be included in its specification. It is likely that the presence of such a drift is what lead our predecessors to doubly smooth, rather than the presence of a repeated unit root.

Autoregressive Error Terms in Regression

$$\text{Model } Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + Z_t$$

Autoregressive error $Z_t = \alpha Z_{t-1} + e_t$
 e_t white noise
 Could be higher order AR(p)

Case 1: α known

$$Z_t = Y_t - (\beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt})$$

$$\alpha Z_{t-1} = \alpha Y_{t-1} - (\alpha \beta_0 + \alpha \beta_1 X_{1,t-1} + \dots + \alpha \beta_k X_{k,t-1})$$

subtract

$$Z_t - \alpha Z_{t-1} = Y_t - \alpha Y_{t-1} - (\beta_0 - \alpha \beta_0) - \beta_1 (X_{1t} - \alpha X_{1,t-1}) + \dots + \beta_k (X_{kt} - \alpha X_{k,t-1})$$

$$\underline{Y_t - \alpha Y_{t-1}} = \beta_0 \underline{(1 - \alpha)} + \beta_1 \underline{(X_{1t} - \alpha X_{1,t-1})} + \dots + \beta_k \underline{(X_{kt} - \alpha X_{k,t-1})} + e_t$$

"New variables" as underlined.

New regression satisfies OLS assumptions (error is e_t)!
 New regression has same coefficients as old!

What about first observation?

Z_1 independent of e_2, e_3, \dots, e_n

$\sqrt{1-\alpha^2} Z_1$ independent of e_2, e_3, \dots, e_n and $\text{Var}\{\sqrt{1-\alpha^2} Z_1\} = \sigma^2$
 $\sqrt{1-\alpha^2} Y_1 = \beta_0 \sqrt{1-\alpha^2} + \beta_1 X_{11} \sqrt{1-\alpha^2} + \dots + \beta_k X_{k1} \sqrt{1-\alpha^2} + \sqrt{1-\alpha^2} Z_1$

| "dependent" | <----- | -"independent"- | -- | -----> |
|-------------------------|---------------------|------------------------------|---------|------------------------------|
| $\sqrt{1-\alpha^2} Y_1$ | $\sqrt{1-\alpha^2}$ | $X_{1,1} \sqrt{1-\alpha^2}$ | \dots | $X_{k,1} \sqrt{1-\alpha^2}$ |
| $Y_2 - \alpha Y_1$ | $1 - \alpha$ | $X_{1,2} - \alpha X_{1,1}$ | | $X_{k,2} - \alpha X_{k,1}$ |
| $Y_3 - \alpha Y_2$ | $1 - \alpha$ | $X_{1,3} - \alpha X_{1,2}$ | | $X_{k,3} - \alpha X_{k,2}$ |
| \vdots | \vdots | \vdots | | \vdots |
| $Y_n - \alpha Y_{n-1}$ | $1 - \alpha$ | $X_{1,n} - \alpha X_{1,n-1}$ | | $X_{k,n} - \alpha X_{k,n-1}$ |

Case 1A: α known, order 2

$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + e_t$

$V_1 = c Z_1$, where $c = \sigma / \sqrt{\gamma(0)}$, has variance σ^2 .

$V_2 = a Z_2 + b Z_1$ has variance $(a^2 + b^2)\gamma(0) + 2ab \gamma(1)$. Set this = σ^2 .

$E\{Z_1 V_2\} = a \gamma(1) + b \gamma(0)$. Set this = 0.

2 equations in 2 unknowns - solve for a,b.

$V_3 = e_3 = Z_3 - \alpha_1 Z_2 - \alpha_2 Z_1$ independent of Z_1, Z_2 ; variance is σ^2 .
 etc.

New **Y** column =
$$\begin{pmatrix} c & 0 & 0 & \dots & 0 \\ b & a & 0 & \dots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}$$

Same for intercept, Xs. Regress new **Y** on new **X** matrix.

Note: This is just an algorithm for doing GLS. The matrix

$$C' C = \begin{pmatrix} c & 0 & 0 & \dots & 0 \\ b & a & 0 & \dots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix}' \begin{pmatrix} c & 0 & 0 & \dots & 0 \\ b & a & 0 & \dots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix}$$

is inverse of $\mathbf{\Gamma} / \sigma^2$ where $\mathbf{\Gamma}$ is variance-covariance matrix of Z_t 's.

Thus $(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{G}^{-1}\mathbf{Y}) = (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{Y})$ which is why regressing $\mathbf{C}\mathbf{Y}$ on $\mathbf{C}\mathbf{X}$ delivers the desired matrix.

Case 2: AR() parameters unknown.

Plan:

1. OLS regression \mathbf{Y} on $\mathbf{X} \Rightarrow$ residuals
2. Fit AR(p) to residuals
3. Apply case 1 algorithm with Estimated AR parameters (EGLS)
4. Use the EGLS estimates to compute new residuals and iterate.

Fuller (Sec. 9.1 and 9.3) gives rather mild conditions under which the above scheme is asymptotically justified (i.e. the transformed regression statistics are asymptotically multivariate normal etc.) when the error series is stationary.

Note 1: If $\alpha = 1$ (residuals are random walk) then the whole model can be specified in differences. Note that we lose the first observation and the ability to estimate the intercept. Nevertheless, computing $Y_t - Y_{t-1}$ as a "new Y" and similarly for the X's is seen to give an equation in transformed variables that has the same β 's as the original equation and has white noise errors.

Note 2: In assessing the contribution of our X's in predicting Y (especially if it is expensive to collect the X information) we note that $\hat{Y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t+1} + \dots + \hat{\beta}_k X_{k,t+1} + \hat{Z}_{t+1}$ where \hat{Z}_{t+1} would be computed from recent residuals and the estimated AR(p) model. Thus if we divide the error sum of squares (errors from these predictions) by the total sum of squares then subtract from 1, we get a "Total R^2 " much of which might come from the momentum in the Z's, especially if they form a nearly nonstationary process. SAS PROC AUTOREG will also produce a "Regression R^2 " which is the R^2 from the regression on the transformed variables and is a better measure of how much you are getting from just the X's.

Note 3: Durbin and Watson suggested the statistic $d = \frac{\sum_1^n (\hat{Z}_t - \hat{Z}_{t-1})^2}{2 \sum_1^n \hat{Z}_t^2}$ as a measure of

autocorrelation. If the Z's are uncorrelated, and we divide numerator and denominator by n, we see that dw converges to 2, so a dw near 2 implies lack of autocorrelation at lag 1. $0 < dw < 2$ implies positive lag 1 autocorrelation and $2 < dw < 4$ implies negative. The dw theory does not allow lagged Ys in the model.

Note 4: The regression in the table above is a nonlinear function of α and the $k+1$ β 's. As such it can be estimated by a general nonlinear likelihood maximization scheme. That is, we can iterate on all the parameters simultaneously rather than iterating on the β s and α separately.

- (1) Run OLS on regression
 Estimators: Unbiased (because $E\{Z\}=0$)
 Consistent (converge to true β s)
- (2) Regress residuals on lagged residuals to ESTIMATE AR coefficients
- (3) Proceed as though estimated AR = true AR.

Example 1: NC Retail Sales (quarterly) updated in size.

Seasonal (qtr 1 low, qtr 4 high) Use dummy variables

Trend: Use t t^2 t^3

Autocorrelation: Try 5 lags (more than a year) and let SAS omit insignificant ones.

```

Data NCSALES;
input qsales t t2 s1 s2 s3 s4 date :yyq6.;
if t2 = . then t2=t*t;
t3=t*t2;
qtr=qtr(date);
format date YYQ4.;
datalines;
    9485.68      2      4      1      0      0      0      1983Q1
    11164.09     3      9      0      1      0      0      1983Q2
    11797.33     4     16      0      0      1      0      1983Q3
    12459.68     5     25      0      0      0      1      1983Q4
    11176.09     6     36      1      0      0      0      1984Q1
    12854.22     7     49      0      1      0      0      1984Q2
                (more data here)
    21130.73    48    2304      0      0      1      0      1994Q3
    22657.68    49    2401      0      0      0      1      1994Q4
      .         50    2500      1      0      0      0      1995Q1
      .         51    2601      0      1      0      0      1995Q2
      .         52    2704      0      0      1      0      1995Q3
      .         53    2809      0      0      0      1      1995Q4
;
    
```

```
proc reg; model qsales = t S1 S2 S3/dwprob;
output out=out1 p=pred lcl=l ucl=u; run;
```

| Parameter Estimates | | | | | |
|---------------------|----|--------------------|----------------|---------|---------|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > t |
| Intercept | 1 | 9556.93236 | 379.69796 | 25.17 | <.0001 |
| t | 1 | 612.17585 | 57.41516 | 10.66 | <.0001 |
| t2 | 1 | -21.51050 | 2.57455 | -8.36 | <.0001 |
| t3 | 1 | 0.28845 | 0.03324 | 8.68 | <.0001 |
| s1 | 1 | -1644.06586 | 196.35286 | -8.37 | <.0001 |
| s2 | 1 | -23.23254 | 195.28585 | -0.12 | 0.9059 |
| s3 | 1 | -124.05869 | 194.64734 | -0.64 | 0.5274 |

| The REG Procedure | |
|----------------------------|--------|
| Model: MODEL1 | |
| Dependent Variable: qsales | |
| Durbin-Watson D | 0.584 |
| Pr < DW | <.0001 |
| Pr > DW | 1.0000 |
| Number of Observations | 48 |
| 1st Order Autocorrelation | 0.662 |

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

```
proc autoreg data=NCSales;
  model qsales = t t2 t3 S1 S2 S3 / nlag=5 backstep dwprob;
run;
```

| The AUTOREG Procedure | | | | | |
|----------------------------------|------------|----------------|-----------|--|--|
| Dependent Variable | | qsales | | | |
| Ordinary Least Squares Estimates | | | | | |
| SSE | 9299521.8 | DFE | 41 | | |
| MSE | 226818 | Root MSE | 476.25372 | | |
| SBC | 747.681588 | AIC | 734.58318 | | |
| Regress R-Square | 0.9737 | Total R-Square | 0.9737 | | |
| Durbin-Watson | 0.5836 | Pr < DW | <.0001 | | |
| Pr > DW | 1.0000 | | | | |

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

| Variable | DF | Estimate | Standard Error | t Value | Approx Pr > t |
|-----------|----|-----------|----------------|---------|----------------|
| Intercept | 1 | 9557 | 379.6980 | 25.17 | <.0001 |
| t | 1 | 612.1759 | 57.4152 | 10.66 | <.0001 |
| t2 | 1 | -21.5105 | 2.5745 | -8.36 | <.0001 |
| t3 | 1 | 0.2884 | 0.0332 | 8.68 | <.0001 |
| s1 | 1 | -1644 | 196.3529 | -8.37 | <.0001 |
| s2 | 1 | -23.2325 | 195.2859 | -0.12 | 0.9059 |
| s3 | 1 | -124.0587 | 194.6473 | -0.64 | 0.5274 |

Estimates of Autocorrelations

| Lag | Covariance | Correlation | |
|-----|------------|-------------|-------|
| 0 | 193740 | 1.000000 | ***** |
| 1 | 128334 | 0.662404 | ***** |
| 2 | 109532 | 0.565354 | ***** |
| 3 | 88836.7 | 0.458535 | ***** |
| 4 | 48426.2 | 0.249955 | ***** |
| 5 | 27365.1 | 0.141246 | *** |

Note: First part same as OLS. Shows strong autocorrelation.

The AUTOREG Procedure

Backward Elimination of Autoregressive Terms

| Lag | Estimate | t Value | Pr > t |
|-----|-----------|---------|---------|
| 5 | 0.074212 | 0.45 | 0.6579 |
| 3 | -0.152076 | -0.85 | 0.4014 |
| 4 | 0.168893 | 1.19 | 0.2425 |
| 2 | -0.225536 | -1.45 | 0.1562 |

Preliminary MSE 108731

Estimates of Autoregressive Parameters

| Lag | Coefficient | Standard Error | t Value |
|-----|-------------|----------------|---------|
| 1 | -0.662404 | 0.118451 | -5.59 |

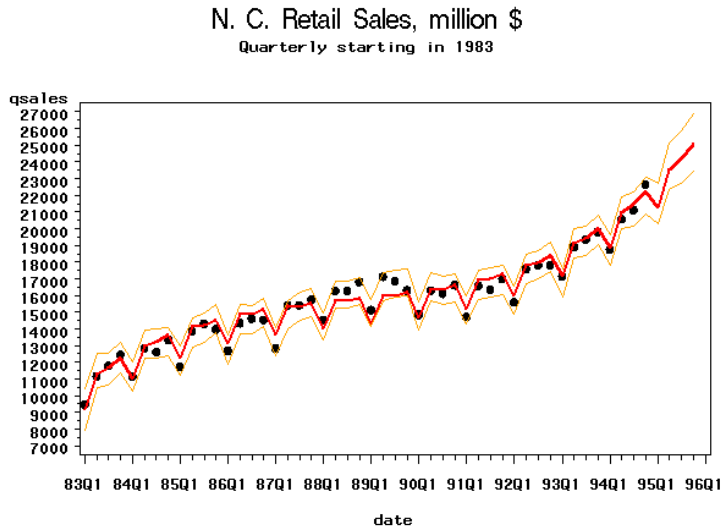
Yule-Walker Estimates

| | | | |
|------------------|------------|----------------|------------|
| SSE | 4736903.38 | DFE | 40 |
| MSE | 118423 | Root MSE | 344.12583 |
| SBC | 719.750619 | AIC | 704.781011 |
| Regress R-Square | 0.9558 | Total R-Square | 0.9866 |
| Durbin-Watson | 1.9768 | Pr < DW | 0.3409 |
| Pr > DW | 0.6591 | | |

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

| Variable | DF | Estimate | Standard Error | t Value | Approx Pr > t |
|-----------|----|-----------|----------------|---------|----------------|
| Intercept | 1 | 9668 | 558.1976 | 17.32 | <.0001 |
| t | 1 | 613.2735 | 92.5406 | 6.63 | <.0001 |
| t2 | 1 | -22.3063 | 4.2045 | -5.31 | <.0001 |
| t3 | 1 | 0.3061 | 0.0542 | 5.65 | <.0001 |
| s1 | 1 | -1620 | 105.4686 | -15.36 | <.0001 |
| s2 | 1 | -11.0318 | 118.3319 | -0.09 | 0.9262 |
| s3 | 1 | -119.7767 | 102.4078 | -1.17 | 0.2491 |

Tests are now valid (asymptotically).



Example 2: NCSU energy usage

Variables:

Demand: Energy used on campus

Class: Classes held that day (dummy variable)

Work: All work days (includes classes)

Temperature

Weekday dummies (base=Wednesday). These add little to R^2

```

/* -----
| NCSU campuswide energy consumption 1980-81 school year. |
| Indicators for workdays and class days.                 |
----- */

proc format; value dow 1="Sun" 2="Mon" 3="Tues" 4="Wed"
               5="Thurs" 6="Fri" 7="Sat";
options pagesize=55 ls=64;

data energy;
input DAY TEMP DEMAND T WORK CLASS DATE :date7. Y 67 ;
      put @20 date;
TEMPSQ=(TEMP-65)**2;

*** compare all days to Wednesday ***;
SAT   = (T=7);   SUN   = (T=1);
MON   = (T=2);   TUES  = (T=3);
THURS = (T=5);   FRI   = (T=6);
      format date date7.;   WC = work+class;
      cval="green";

```

```
if WC=0 then cval="red";
if WC=1 then cval = "cyan";
```

```
datalines;
      1      83      8217      1      0      0      01JUL79      2
      2      87      12545      2      1      1      02JUL79      1
      3      85      12649      3      1      1      03JUL79      1
```

(more data)

```
      30      88      13789      2      1      1      30JUN80      1
```

```
;
```

```
PROC GPLOT data=energy; PLOT DEMAND*TEMP=WC;
```

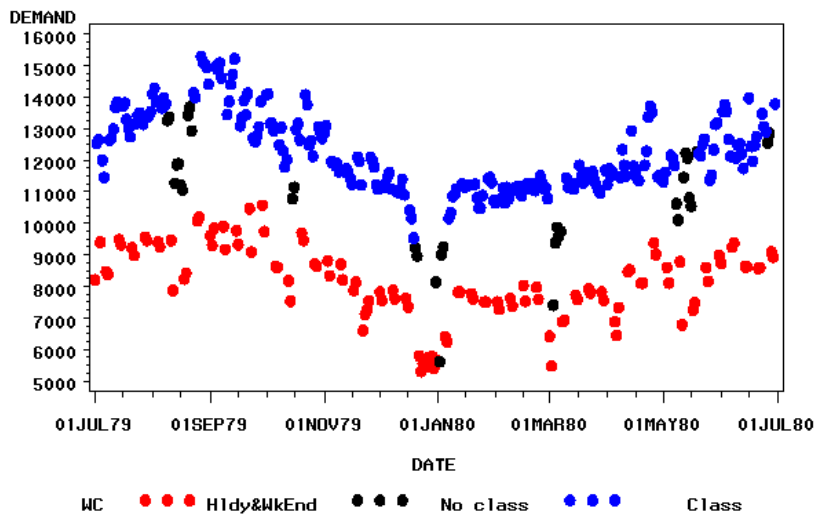
```
  title 'NCSU Energy Demand';
  title2 c=red "Off " c=cyan " Work, no Class " c=green " Class ";
  Symbol1 v=dot i=none c=red;
  Symbol2 v=dot i=none c=cyan;;
  symbol3 v=dot i=none c=green;
```

```
PROC GPLOT data=energy; PLOT DEMAND*T=WC ;
      format t dow.;
```

```
PROC AUTOREG data=energy; MODEL DEMAND = TEMP TEMPSQ CLASS WORK
      SAT--FRI /NLAG=15 BACKSTEP;
      ** day of week adds about .002 to R**2 ***;
run;
```

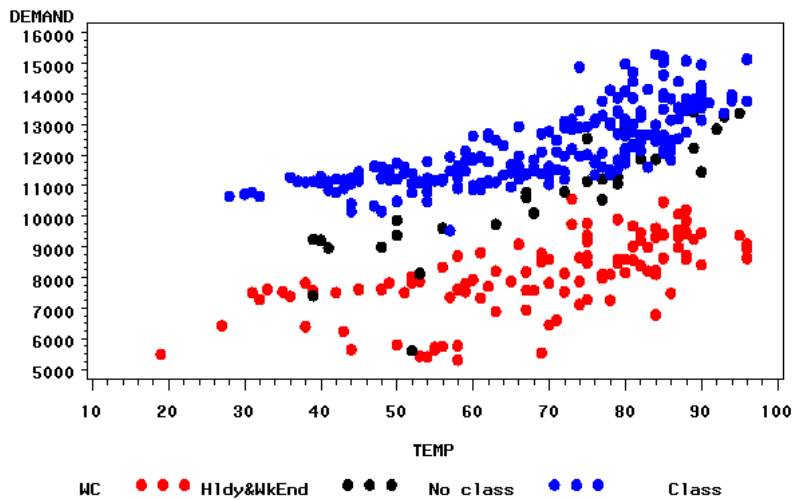
A simple plot of energy usage versus temperature using as a symbol the sum of the class and workday seasonal dummies (0-> not a work day, 1-> work day without classes, 2-> a class day).

Energy demand at NCSU



CLASS DAY
WORK DAY - NO CLASSES
HOLIDAY or WEEKEND

Energy demand at NCSU



Autoreg Procedure

Dependent Variable = DEMAND

Ordinary Least Squares Estimates

| | | | |
|---------------|----------|-----------|----------|
| SSE | 2.5387E8 | DFE | 361 |
| MSE | 703238.1 | Root MSE | 838.5929 |
| SBC | 5990.765 | AIC | 5971.252 |
| Reg Rsq | 0.8680 | Total Rsq | 0.8680 |
| Durbin-Watson | 0.6324 | | |

| Variable | DF | B Value | Std Error | t Ratio | Approx Prob |
|-----------|----|-------------|-----------|---------|-------------|
| Intercept | 1 | 3856.671264 | 206.8 | 18.650 | 0.0001 |
| TEMP | 1 | 57.833448 | 2.6720 | 21.644 | 0.0001 |
| TEMPSQ | 1 | 0.824102 | 0.1523 | 5.412 | 0.0001 |
| CLASS | 1 | 1545.669155 | 157.1 | 9.841 | 0.0001 |
| WORK | 1 | 2638.304419 | 165.6 | 15.934 | 0.0001 |

Estimates of Autocorrelations

| Lag | Covariance | Correlation |
|-----|------------|-------------|
| 0 | 693631.1 | 1.000000 |
| 1 | 473562.1 | 0.682729 |
| 2 | 367778.3 | 0.530222 |
| 3 | 308403.7 | 0.444622 |
| 4 | 307188 | 0.442869 |
| 5 | 333956.1 | 0.481461 |
| 6 | 342626.4 | 0.493961 |
| 7 | 365283.9 | 0.526626 |
| 8 | 298846.6 | 0.430844 |
| 9 | 299775.9 | 0.432183 |
| 10 | 252863.5 | 0.364550 |
| 11 | 221802.3 | 0.319770 |
| 12 | 221137.3 | 0.318811 |
| 13 | 220904.3 | 0.318475 |
| 14 | 276792.2 | 0.399048 |
| 15 | 231104.3 | 0.333180 |

Backward Elimination of Autoregressive Terms

| Lag | Estimate | t-Ratio | Prob |
|-----|-----------|---------|--------|
| 2 | 0.002017 | 0.0335 | 0.9733 |
| 12 | -0.005618 | -0.0938 | 0.9253 |
| 3 | 0.004789 | 0.0890 | 0.9291 |
| 10 | 0.054058 | 0.9333 | 0.3513 |
| 4 | -0.060362 | -1.1469 | 0.2522 |
| 6 | -0.079523 | -1.3325 | 0.1836 |
| 13 | 0.057813 | 1.0993 | 0.2724 |
| 11 | 0.076009 | 1.6618 | 0.0974 |
| 15 | 0.098888 | 1.8942 | 0.0590 |

Preliminary MSE = 314640.5

Estimates of the Autoregressive Parameters

| Lag | Coefficient | Std Error | t Ratio |
|-----|-------------|-----------|---------|
| 1 | -0.54064918 | 0.044220 | -12.226 |
| 5 | -0.11926166 | 0.044812 | -2.661 |
| 7 | -0.20036592 | 0.054040 | -3.708 |
| 8 | 0.16453292 | 0.058817 | 2.797 |
| 9 | -0.10386205 | 0.050991 | -2.037 |
| 14 | -0.10107126 | 0.044566 | -2.268 |

Expected Autocorrelations

| Lag | Autocorr |
|-----|----------|
| 0 | 1.0000 |
| 1 | 0.6883 |
| 2 | 0.5369 |
| 3 | 0.4527 |
| 4 | 0.4365 |
| 5 | 0.4771 |
| 6 | 0.4799 |
| 7 | 0.5188 |
| 8 | 0.4278 |
| 9 | 0.4298 |
| 10 | 0.4072 |
| 11 | 0.3919 |
| 12 | 0.3988 |
| 13 | 0.3992 |
| 14 | 0.4427 |

| Autoreg Procedure | | | | | |
|---------------------------|----------|-------------|-----------|---------|-------------|
| Yule-Walker Estimates | | | | | |
| SSE | 96233172 | DFE | 355 | | |
| MSE | 271079.4 | Root MSE | 520.6528 | | |
| SBC | 5672.801 | AIC | 5629.872 | | |
| Reg Rsq | 0.8881 | Total Rsq | 0.9500 | | |
| Durbin-Watson | 1.7564 | | | | |
| Variable | DF | B Value | Std Error | t Ratio | Approx Prob |
| Intercept | 1 | 5857.458298 | 374.1 | 15.659 | 0.0001 |
| TEMP | 1 | 31.686338 | 3.6911 | 8.585 | 0.0001 |
| TEMPSQ | 1 | 0.687270 | 0.1232 | 5.579 | 0.0001 |
| CLASS | 1 | 1185.053609 | 119.9 | 9.882 | 0.0001 |
| WORK | 1 | 2763.301101 | 123.7 | 22.345 | 0.0001 |
| Expected Autocorrelations | | | | | |
| Lag | Autocorr | | | | |
| 0 | 1.0000 | | | | |
| 1 | 0.6883 | | | | |
| 2 | 0.5369 | | | | |
| 3 | 0.4527 | | | | |
| 4 | 0.4365 | | | | |
| 5 | 0.4771 | | | | |
| 6 | 0.4799 | | | | |
| 7 | 0.5188 | | | | |
| 8 | 0.4278 | | | | |
| 9 | 0.4298 | | | | |
| 10 | 0.4072 | | | | |
| 11 | 0.3919 | | | | |
| 12 | 0.3988 | | | | |
| 13 | 0.3992 | | | | |
| 14 | 0.4427 | | | | |

Note: AUTOREG assumes model is $Z_t + \alpha_1 Z_{t-1} + \dots + \alpha_p Z_{t-p} = e_t$. Parameter estimates have opposite signs from Box-Jenkins ARIMA parameters.

Note: AUTOREG is the only SAS/ETS procedure that includes a variance for the regression parameters in its prediction interval computations. Adds $\mathbf{X}_{t+j} \hat{\boldsymbol{\beta}}$ variance to \hat{Z}_{t+j} prediction error variance, but future \mathbf{X} matrix row, \mathbf{X}_{t+j} treated as a *constant* but is usually really a forecast in practice. Demos: NCSales.sas and Energy.sas

Transfer Functions:

Think about ARMA

$$(Y_t - \mu) = \alpha (Y_{t-1} - \mu) + e_t - \theta e_{t-1}$$

Backshift:

$$(1 - \alpha B) (Y_t - \mu) = (1 - \theta B) e_t$$

$$(Y_t - \mu) = \frac{(1 - \theta B)}{(1 - \alpha B)} e_t = \frac{\text{MA backshift}}{\text{AR backshift}} e_t$$

How do we "identify" ? Covariances! (ACF)

| | | | | | |
|---------------------------------|---|----|----|----|----|
| h | 0 | 1 | 2 | 3 | 4 |
| $\rho(h) = \gamma(h)/\gamma(0)$ | 1 | .8 | .4 | .2 | .1 |

We know that it is ARMA(1,1) from this pattern and $\alpha = .5$. Can also solve for θ .

Definition: Cross covariance.

$$E\{ X_t \} = \mu_X \quad E\{ Y_t \} = \mu_Y$$

$$\gamma_{XY}(h) = E\{ (X_t - \mu_X)(Y_{t+h} - \mu_Y) \}$$

Example: $X_t = .8 X_{t-1} + e_t$, $Y_t = 10 X_{t-1} + .5 Y_{t-1} + e_t + 0.8 e_{t-1}$
 e_t and e_{t-1} independent $N(0,1)$

multiply by X_t , take $E\{ \}$

$$\gamma_{XY}(0) = E\{ X_t Y_t \} = 10 \gamma_{XX}(1) + .5 \gamma_{XY}(-1) + 0 + 0.8 (0.8)$$

by X_{t-1}

$$\gamma_{XY}(1) = E\{ X_{t-1} Y_t \} = 10 \gamma_{XX}(0) + .5 \gamma_{XY}(0) + 0 + 0.8 (1)$$

etc.

Better way:

$$(1 - .5B) Y_t = 10 X_{t-1} + e_t + .8 e_{t-1}$$

$$Y_t = 1/(1 - .5B) [10 X_{t-1} + e_t + .8 e_{t-1}]$$

$$= (1 + .5B + .25 B^2 + \dots) [10 X_{t-1} + e_t + .8 e_{t-1}]$$

so

$$E\{ X_{t-1} Y_t \} = 10 \gamma_{XX}(1) + 5 \gamma_{XX}(0) + 2.5 \gamma_{XX}(1) + 1.25 \gamma_{XX}(2) + \dots$$

$$+ .5E\{ X_{t-1} e_{t-1} \} + .25 E\{ X_{t-1} e_{t-2} \} + \dots$$

where $\gamma_{XX}(j) = \frac{1}{0.36} .8^j$ and $E\{ X_t e_{t-j} \} = .8^j$

Points:

Relationship between model and $\gamma_{XY}(h)$ is 1-1 but complicated.

Why is it so messy? Because X is autocorrelated!

Look at

$$X_t = e_t,$$

$$Y_t = 10 X_{t-1} + .5 Y_{t-1} + \epsilon_t + .8 e_{t-1} (= 10.8 X_{t-1} + .5 Y_{t-1} + \epsilon_t)$$

ϵ_t and e_t independent $N(0,1)$.

This time

$$Y_t = 1/(1-.5B)[10 X_{t-1} + \epsilon_t + .8 e_{t-1}]$$

$$= (1 + .5B + .25 B^2 + \dots) [10.8 X_{t-1} + \epsilon_t]$$

$$= 10.8 X_{t-1} + 5.4 X_{t-2} + 2.7 X_{t-3} + 1.35 X_{t-4} + \dots + (1 + .5B + .25 B^2 + \dots)\epsilon_t$$

Multiply by X_{t-j} and take $E\{ \}$. $E\{X_{t-j}Y_t\} = 0, 10.8\gamma_{XX}(0), 5.4\gamma_{XX}(0), 2.7\gamma_{XX}(0)$ etc. for $j=0,1,2,3$, etc. These numbers are proportional to the coefficients on lagged X . They show the pure delay of 1 and then exponential decay (like $AR(1)$).

Look at Y_t . What if we ignore the error ϵ_t in Y and give "pulse" input X : $X = 0, 0, 0, 0, 0, 1, 0, 0, 0, 0$. What is Y response?

$$X = 0, 0, 0, 0, 0, 1, 0, 0, 0, 0 \text{ etc.}$$

$$Y = 0, 0, 0, 0, 0, 10.8, 5.4, 2.7, 1.35 \text{ etc.}$$

so these numbers are called "impulse response" weights or impulse response function.

When X is white noise, $\gamma_{XY}(h)$ is proportional to impulse response function and this in turn shows the backshift operators relating Y to X . In our example,

$AR(1)$ -type behavior \Rightarrow 1 lag in denominator

$$Y_t = \frac{1}{1-\alpha B} X_{t-1} + \text{noise} = \text{transfer function} + \text{noise}.$$

Divide cross-covariance by $\sqrt{\text{product of variances}}$ to make it cross *correlation*
Practice:

Suppose X is white noise and you see these cross correlations:

| h | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|----------|----|----|----|----|----|-----|------|-------|
| series 1 | 0 | 0 | 0 | .8 | .4 | 0 | 0 | 0 |
| series 2 | 0 | .9 | .4 | .2 | .1 | .05 | .025 | .0125 |
| series 3 | 0 | 0 | .3 | .8 | .4 | .2 | .1 | .05 |
| series 4 | 0 | .8 | 0 | 0 | 0 | .3 | 0 | 0 |

Series 1: $Y_t = \left(\frac{\quad}{\quad} \right) X_{t- \quad} + \text{noise}$

Series 2: $Y_t = \left(\frac{\quad}{\quad} \right) X_{t- \quad} + \text{noise}$

Example:

In SAS to fit series 1 you would say

```
PROC ARIMA; I VAR=Y CROSSCOR=(X);
E INPUT = ( 2$(1) X) PLOT;
```

For series 2:

```
PROC ARIMA; I VAR=Y CROSSCOR=(X);
E INPUT = ( (1)/(1) X);
```

What if X is *not* white noise?

Simple example: $Y_t = 3 X_{t-3} + e_t$ but $X_t = .8 X_{t-1} + \epsilon_t$

In this case $\gamma_{XY}(h) = 3 \gamma_{XX}(h-3)$ which is nonzero at all lags.

The true relationship (lag 3) has been "*smearred*" out across all lags in $\gamma_{XY}(h)$.

How to fix this? If we know $X_t = .8 X_{t-1} + \epsilon_t$ then we can do this:

$$Y_t = 3 X_{t-3} + e_t$$

$$.8 Y_{t-1} = 3(.8 X_{t-4}) + .8e_{t-1}$$

subtract:

$$(Y_t - .8 Y_{t-1}) = 3 (X_{t-3} - .8 X_{t-4}) + e_t - .8e_{t-1}$$

$$(Y_t - .8 Y_{t-1}) = 3 (\epsilon_{t-3}) + e_t - .8e_{t-1}$$

so we have 2 *new* variables $(Y_t - .8 Y_{t-1})$ and $(X_t - .8 X_{t-1})$ whose relationship is exactly the one we are after *and* $(X_t - .8 X_{t-1}) = \epsilon_t$ is white noise.

Plan:

Model X_t

Use X_t model to "filter" X and Y. This is *prewhitening*

Cross correlate the prewhitened variables.

After diagnosing the structure, fit on original variables

Plot ACF etc. of residuals

Diagnose noise structure

Fit transfer function plus noise

Diagnosing:

Look at cross correlations like autocorrelations

AR behavior corresponds to denominator lags

MA behavior corresponds to numerator lags

Some alternative ideas for model building:

You can always use trial and error - i.e. the prewhitening and crosscorrelations

are only aids in diagnosing structure and can be bypassed.

You can fit a long set of lags: $INPUT = (1,2,3,4,5,6,7,8) X$ and try to diagnose the coefficient pattern without prewhitening. Here you are counting on the coefficients declining rather sharply with lag number.

Example: Flows of the Neuse River.

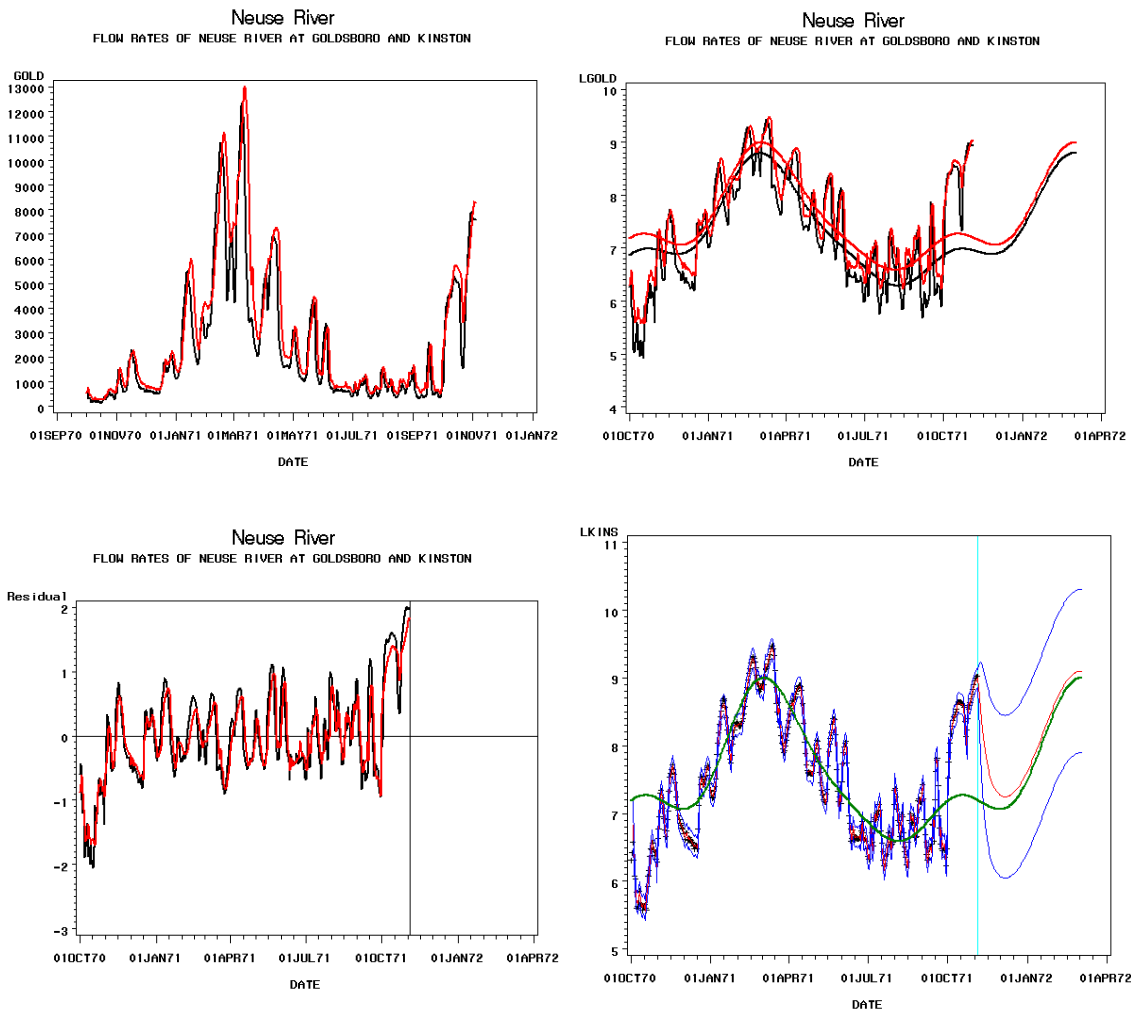
Daily flows, log transformed.
 Deseasonalized using sinusoids.

$G(t)$ = Flow upstream at Goldsboro (logs, deseasonalized)

$K(t)$ = Flow downstream at Kinston (logs, deseasonalized)

$G(t) \rightarrow K(t)$

Demo Program Neuse.sas gives nice graphs.



Summary of Neuse.sas program:

(1) Crosscorrelate K(t) with G(t-L) without prewhitening

| Crosscorrelations | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------------|------------|-------------|----|---|---|---|---|---|---|---|---|---|---|-------|---|---|---|---|---|---|---|---|---|--|
| Lag | Covariance | Correlation | -1 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 1 | |
| -8 | 0.164394 | 0.35074 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -7 | 0.183985 | 0.39254 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -6 | 0.208450 | 0.44473 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -5 | 0.239274 | 0.51050 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -4 | 0.275069 | 0.58687 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -3 | 0.316023 | 0.67424 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -2 | 0.358871 | 0.76566 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| -1 | 0.397895 | 0.84892 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 0 | 0.432825 | 0.92344 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 1 | 0.447712 | 0.95520 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 2 | 0.436797 | 0.93192 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 3 | 0.406182 | 0.86660 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 4 | 0.366343 | 0.78160 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 5 | 0.322783 | 0.68867 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 6 | 0.280843 | 0.59919 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 7 | 0.242964 | 0.51837 | | | | | | | | | | | . | ***** | | | | | | | | | | |
| 8 | 0.209769 | 0.44755 | | | | | | | | | | | . | ***** | | | | | | | | | | |

"." marks two standard errors

Confusing! Correlations at negative lags (river flowing backwards ????)

Lags of 8 days too much! (stations only 50 kilometers apart)

Why did this happen? (G(t) highly autocorrelated - failed to prewhiten)

(2) Fit Model to G(t), use to prewhiten K(t) and G(t)

New crosscorrelations:



| | | | | | | | | | |
|----|------------|---------|--|----|--|-------|---|--|--|
| -8 | 0.00094314 | 0.03433 | | . | | * | . | | |
| -7 | -0.0000525 | -.00191 | | . | | . | . | | |
| -6 | -0.0010276 | -.03740 | | .* | | . | . | | |
| -5 | 0.00051780 | 0.01885 | | . | | . | . | | |
| -4 | -0.0015174 | -.05523 | | .* | | . | . | | |
| -3 | 0.0013477 | 0.04905 | | . | | * | . | | |
| -2 | 0.0057804 | 0.21040 | | . | | **** | . | | |
| -1 | -0.0015216 | -.05538 | | .* | | . | . | | |
| 0 | 0.00029230 | 0.01064 | | . | | . | . | | |
| 1 | 0.016915 | 0.61568 | | . | | ***** | . | | |
| 2 | 0.0088699 | 0.32285 | | . | | ***** | . | | |
| 3 | 0.00023157 | 0.00843 | | . | | . | . | | |
| 4 | 0.0018391 | 0.06694 | | . | | * | . | | |
| 5 | 0.0012101 | 0.04405 | | . | | * | . | | |
| 6 | 0.00029431 | 0.01071 | | . | | . | . | | |
| 7 | 0.00013617 | 0.00496 | | . | | . | . | | |
| 8 | -0.0005137 | -.01870 | | . | | . | . | | |

"." marks two standard errors

Class exercise - This tells us what? Estimate input = (_____ Goldsboro) ;

(3) Now that the transfer relationship is identified, go back and fit the model, plotting the error ACF, PACF, IACF. Diagnose 2 AR lags, 4 MA lags.

(4) Refit the transfer function model with error autocorrelation p=2, q=4.

| Maximum Likelihood Estimation | | | | | | | | |
|-------------------------------|----------|----------------|---------|----------------|-----|----------|-------|--|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag | Variable | Shift | |
| MU | 0.01589 | 0.13794 | 0.12 | 0.9083 | 0 | rkins | 0 | |
| MA1,1 | 0.18902 | 0.13245 | 1.43 | 0.1535 | 1 | rkins | 0 | |
| MA1,2 | 0.30724 | 0.05528 | 5.56 | <.0001 | 2 | rkins | 0 | |
| MA1,3 | 0.26405 | 0.06783 | 3.89 | <.0001 | 3 | rkins | 0 | |
| MA1,4 | 0.16633 | 0.07897 | 2.11 | 0.0352 | 4 | rkins | 0 | |
| AR1,1 | 1.43422 | 0.12920 | 11.10 | <.0001 | 1 | rkins | 0 | |
| AR1,2 | -0.43700 | 0.12621 | -3.46 | 0.0005 | 2 | rkins | 0 | |
| NUM1 | 0.50260 | 0.01888 | 26.63 | <.0001 | 0 | rgold | 1 | |
| NUM1,1 | -0.27829 | 0.01872 | -14.86 | <.0001 | 1 | rgold | 1 | |

$$K(t) = 0.0 + 0.50 G(t-1) - (-.28) G(t-2) + \text{ARMA}(2,4)$$

(5) Check of fit automatically given

(5A) Check for crosscorrelation of error with lagged inputs (H0: transfer part is OK)

Crosscorrelation Check of Residuals with Input rgold

| To Lag | Chi-Square | DF | Pr > ChiSq | -----Crosscorrelations----- | | | | | |
|--------|------------|----|------------|-----------------------------|--------|--------|--------|--------|--------|
| 5 | 6.88 | 5 | 0.2296 | -0.051 | -0.082 | -0.016 | 0.051 | 0.060 | 0.039 |
| 11 | 14.83 | 11 | 0.1903 | 0.087 | 0.031 | 0.090 | 0.022 | -0.046 | -0.028 |
| 17 | 27.10 | 17 | 0.0566 | 0.015 | -0.023 | 0.034 | 0.137 | -0.052 | 0.087 |
| 23 | 28.88 | 23 | 0.1842 | 0.043 | -0.039 | -0.001 | 0.014 | -0.010 | 0.029 |
| 29 | 33.29 | 29 | 0.2663 | -0.034 | -0.062 | -0.055 | -0.032 | -0.038 | -0.026 |
| 35 | 40.00 | 35 | 0.2578 | -0.006 | -0.037 | -0.014 | -0.107 | 0.024 | 0.057 |
| 41 | 47.07 | 41 | 0.2380 | 0.083 | 0.027 | -0.032 | -0.064 | -0.012 | 0.070 |
| 47 | 54.08 | 47 | 0.2223 | -0.079 | -0.046 | 0.083 | 0.040 | -0.021 | -0.022 |

(5B) Autocorrelation check (Box-Ljung statistics H0: error model is OK)

Note: Get transfer part right first, then model errors.

Autocorrelation Check of Residuals

| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
|--------|------------|----|------------|----------------------------|--------|--------|--------|--------|--------|
| 6 | . | 0 | . | 0.003 | -0.005 | -0.015 | -0.018 | -0.048 | 0.002 |
| 12 | 11.65 | 6 | 0.0703 | -0.028 | -0.077 | 0.113 | -0.000 | 0.067 | -0.040 |
| 18 | 21.23 | 12 | 0.0471 | -0.025 | -0.011 | 0.021 | -0.141 | 0.033 | 0.031 |
| 24 | 30.72 | 18 | 0.0310 | 0.014 | 0.130 | -0.001 | 0.061 | -0.019 | -0.037 |
| 30 | 34.74 | 24 | 0.0723 | 0.005 | -0.027 | -0.032 | -0.074 | -0.008 | -0.045 |
| 36 | 39.28 | 30 | 0.1197 | 0.026 | -0.014 | 0.041 | 0.037 | 0.036 | -0.072 |
| 42 | 40.35 | 36 | 0.2840 | -0.014 | -0.004 | -0.023 | 0.032 | -0.016 | -0.020 |
| 48 | 47.17 | 42 | 0.2695 | 0.022 | 0.104 | 0.046 | 0.028 | 0.029 | -0.016 |

So - model looks OK, not perfect.

Intervention Models:

These are the same as the transfer function models *except* that the X variable is some sort of indicator variable.

Example:

$$Y_t = 100 + .8(Y_{t-1} - 100) + 3(X_t - 5) + 2(X_{t-1} - 5) + \text{noise}$$

Suppose we ignore the noise, X_t has been 5 and Y_t has been 100 for some time. Now X receives a pulse at time 20, say $X_{20} = 6$, then X_t drops back to 5 and stays there. What will happen to Y?

$$Y_{20} = 100 + .8(0) + 3(6-5) + 2(0) = 103$$

$$Y_{21} = 100 + .8(3) + 3(0) + 2(1) = 104.4$$

$$Y_{22} = 100 + .8(4.4) + 3(0) + 2(0) = 103.52$$

Alternatively we write

$$(1 - .8B)(Y_t - 100) = (3 + 2B)(X_t - 5)$$

$$(Y_t - 100) = [(3 + 2B) / (1 - .8B)](X_t - 5) \\ = (3 + 4.4B + 3.52B^2 + 4.4(.8^2)B^3 + \dots)(X_t - 5)$$

which displays the impulse response function.

Fitting is the same as any other transfer function. However it is not possible to "filter" a dummy variable to get white noise so the diagnosis of the transfer function form, $(3 + 2B) / (1 - .8B)$ in our case, is done by just looking at the response of Y near the intervention point.

The intervention could also be a *permanent* change from 0 to 1 so that $X_t = 0$ for $t < 20$ and $X_t = 1$ for $t > 19$. In that case, the response of Y would be partial sums of the above numbers, that is,

3, 7.4, 10.92, etc.

One might notice an arbitrary jump from 3 to 7.4 (indicating a numerator lag) followed by jumps each of which is .8 times its predecessor (indicating one denominator lag with coefficient .8). Of course this is harder when the data are contaminated by noise which will always be the case.

Example: Suppose there is an initial drop, -5, followed by levels 3, 7, 9, 10, 10.5, 10.75 etc. converging to 11. We see an initial drop to -5 then we have $-5 + 8 = 3$ then $3 + 4 = 7$ then $7 + 2 = 9$ then $9 + 1 = 10$ etc. We have $(-5 + 10.5B) / (1 - .5B)$ operating on a level shift dummy variable. This might happen, for example, if a company makes a major change. Its stock initially drops then if the change is a good one, it rises to a higher level than that at which it started. In SAS all we need to know is *how many* numerator and denominator lags are needed. SAS will estimate the coefficients.

```
PROC ARIMA;
  IDENTIFY VAR=Y CROSSCOR=(X) NOPRINT;
  ESTIMATE INPUT = ( (1)/(1) X ) PLOT;
```

Because no model is available for a dummy variable, you must supply its values into the future for your forecasts.

One more example:

Intervention at time 30 (permanent shift from 0 to 1),

$Y=0$ up to time 32. $Y_{33} = 10$, $Y_t = 17$ for $t>33$.

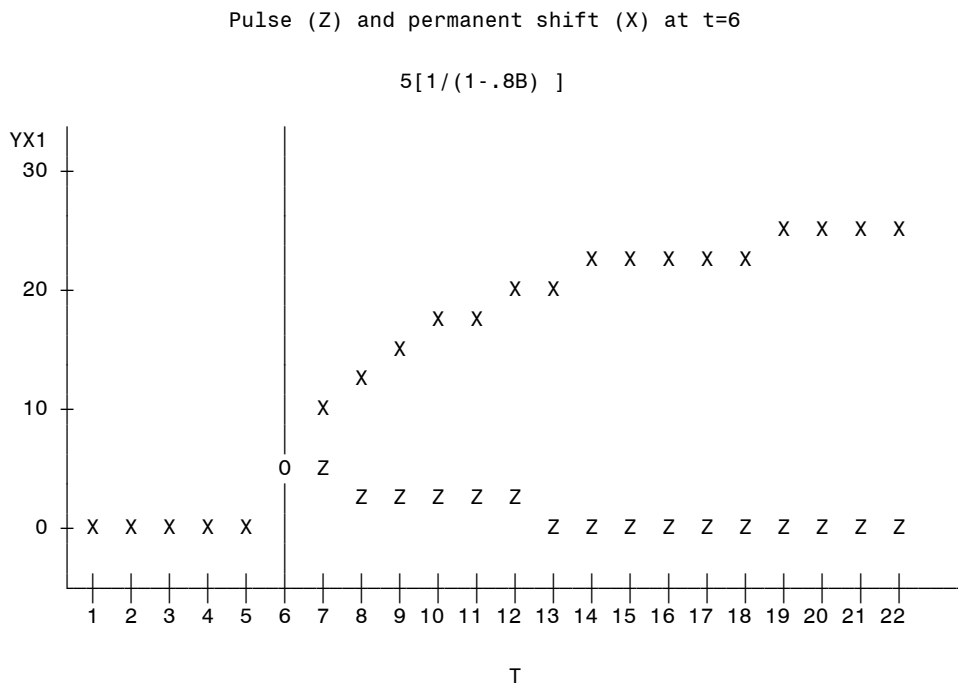
we see that

$$Y_t = 10 X_{t-3} + 7 X_{t-4}$$

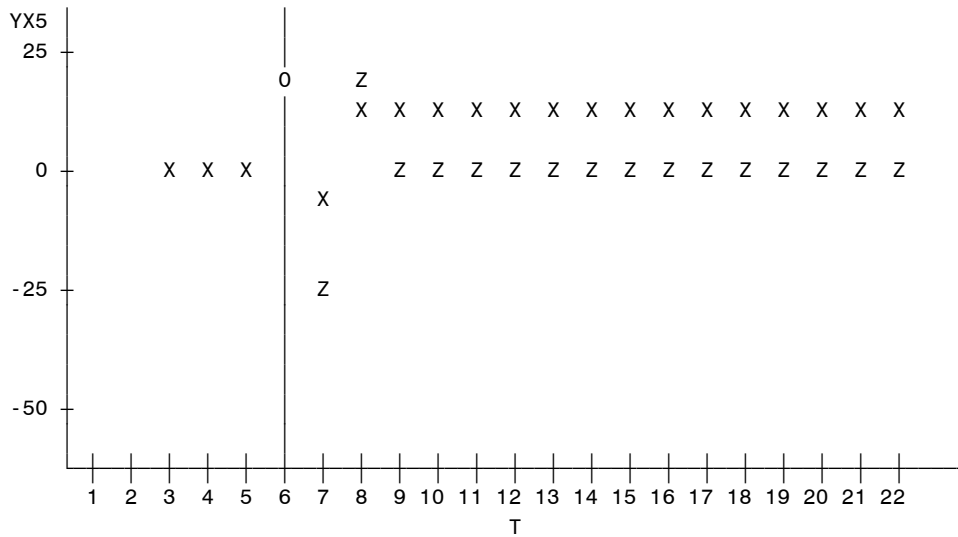
```
ESTIMATE INPUT = ( 3$(1) X );
```

We see that there is some sort of delay in the Y response, possibly due to lack of immediate information about X, then a two period adjustment with the second period Y being completely adjusted to the new level.

Here are some graphs of response to a pulse (Z) or level shift (X) at time 6 for various backshift operators:



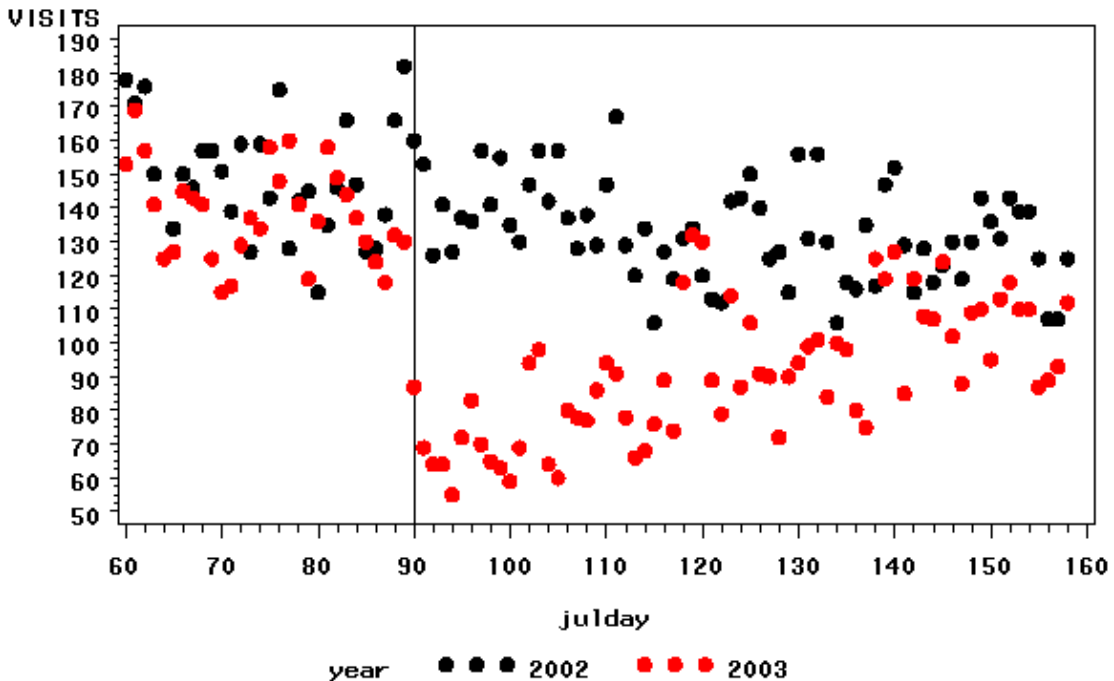
Pulse (Z) and permanent shift (X) at t=6
 $20*(1-1.38 B+.95 B**2)$



Example: We looked at elective visits to the hospital in Toronto in 2002. In 2003 the SARS epidemic broke out and such elective visits dropped off dramatically but eventually returned to the pattern seen prior to SARS. Here is an intervention analysis with a graphically determined date for the SARS pulse effect (this was several days after the first cases appeared).

Hospital elective visits

Assigning 31mar2003 as start of SARS scare



&days is the macro string d1 d2 d3 d4 d5 d6.

julday is Julian day (within the year).

SARS is 0 except for a 1 at the graphically chosen impact date in 2003.

```
proc arima data=SARS2; * where year=3;
identify var=visits
  crosscor = (d_slope julday SARS perm &days year)
  noprint;
estimate input = (julday /(1)SARS &days) p=1 ml;
forecast lead=0 id=date interval=1 out=out2;
```

| Maximum Likelihood Estimation | | | | | | | |
|-------------------------------|-----------|----------------|---------|----------------|-----|----------|-------|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag | Variable | Shift |
| MU | 162.48914 | 5.52380 | 29.42 | <.0001 | 0 | VISITS | 0 |
| AR1,1 | 0.24517 | 0.07130 | 3.44 | 0.0006 | 1 | VISITS | 0 |
| NUM1 | -0.19969 | 0.04913 | -4.06 | <.0001 | 0 | julday | 0 |
| NUM2 | -74.65255 | 5.47064 | -13.65 | <.0001 | 0 | SARS | 0 |
| DEN1,1 | 0.98084 | 0.0030525 | 321.32 | <.0001 | 1 | SARS | 0 |
| NUM3 | 7.87668 | 3.02439 | 2.60 | 0.0092 | 0 | D1 | 0 |
| NUM4 | 0.12568 | 3.37374 | 0.04 | 0.9703 | 0 | D2 | 0 |
| NUM5 | -8.86646 | 3.44707 | -2.57 | 0.0101 | 0 | D3 | 0 |
| NUM6 | -10.70244 | 3.44685 | -3.10 | 0.0019 | 0 | D4 | 0 |
| NUM7 | -15.18572 | 3.37303 | -4.50 | <.0001 | 0 | D5 | 0 |
| NUM8 | -7.23870 | 3.03037 | -2.39 | 0.0169 | 0 | D6 | 0 |
| Constant Estimate | | | | 122.6521 | | | |
| Variance Estimate | | | | 161.5999 | | | |
| Std Error Estimate | | | | 12.71219 | | | |
| AIC | | | | 1571.574 | | | |
| SBC | | | | 1607.689 | | | |
| Number of Residuals | | | | 197 | | | |

$$V_t = 162 - 0.2(\text{day of year}) - \frac{75}{(1-0.98B)}(\text{SARS}) + D(\text{day}) + \frac{1}{(1-0.25B)}e_t$$

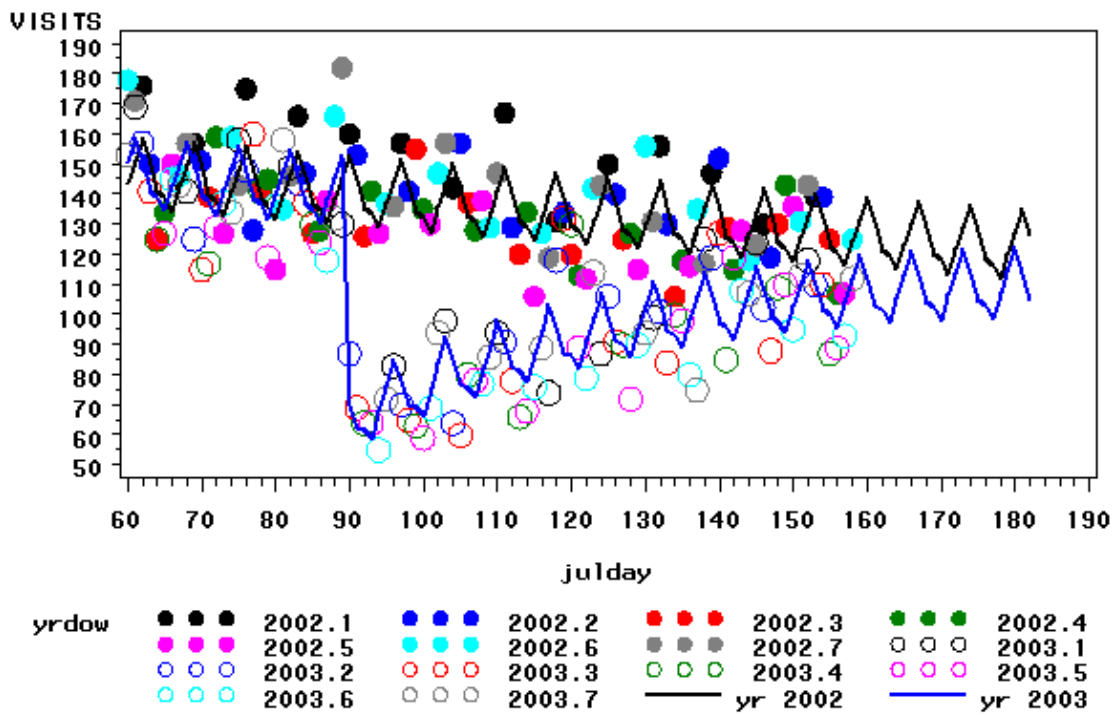
D(day) effects

| Sun | Mon | Tues | Wed | Thurs | Fri | Sat |
|-----|-----|------|-------|-------|------|-----|
| 7.9 | 0.1 | -8.9 | -10.7 | -15.2 | -7.2 | (0) |

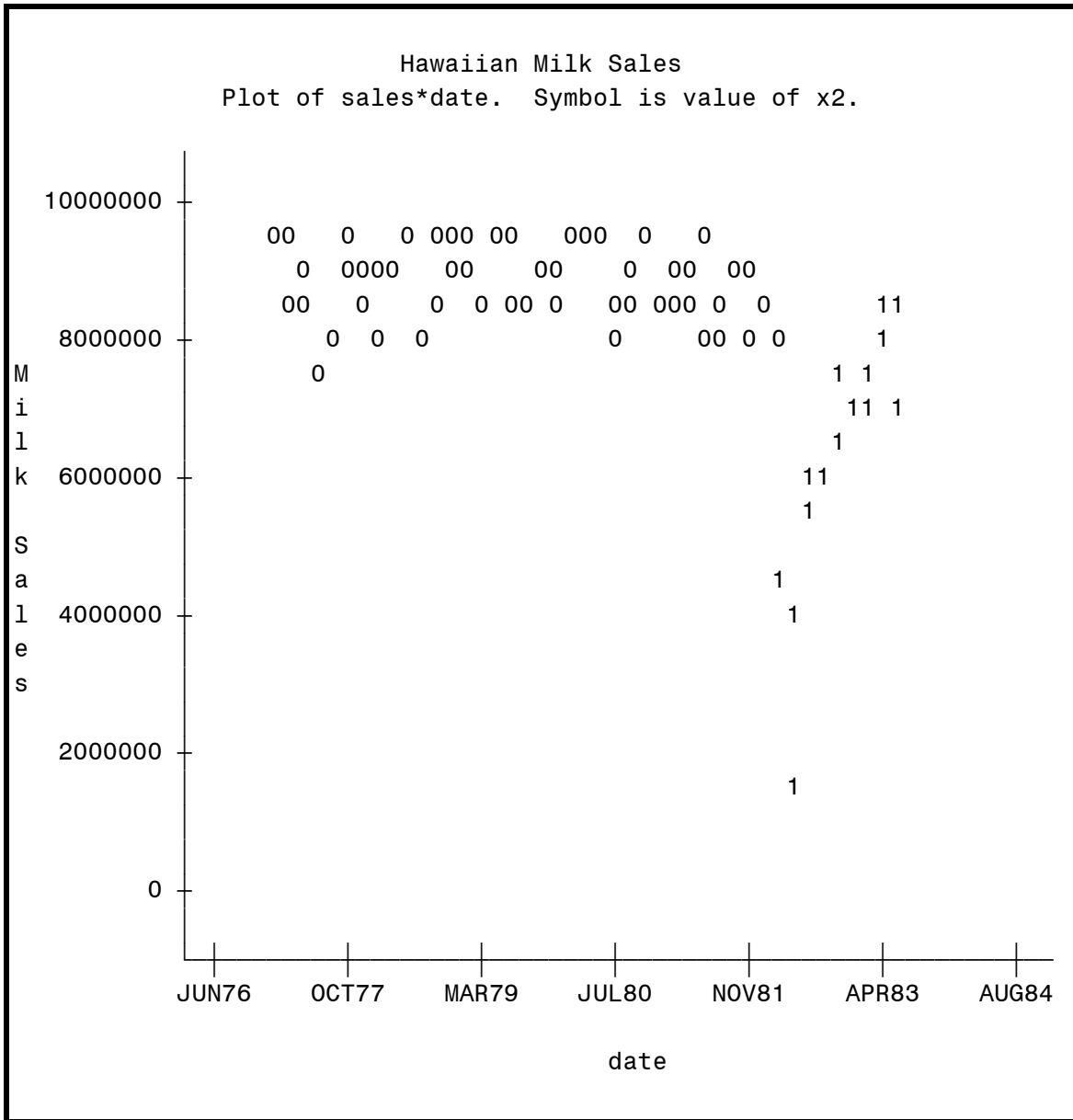
$$\hat{\sigma}^2 = 161.6$$

| Autocorrelation Check of Residuals | | | | | | | | | |
|------------------------------------|------------|----|------------|----------------------------|--------|--------|--------|-------|--------|
| To Lag | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | |
| 6 | 1.92 | 5 | 0.8601 | -0.009 | 0.044 | -0.078 | 0.030 | 0.011 | -0.018 |
| 12 | 5.25 | 11 | 0.9185 | -0.043 | -0.092 | 0.067 | -0.018 | 0.011 | -0.029 |
| 18 | 8.78 | 17 | 0.9469 | 0.045 | -0.055 | -0.013 | 0.005 | 0.060 | -0.086 |
| 24 | 11.30 | 23 | 0.9799 | 0.046 | 0.064 | 0.029 | 0.062 | 0.020 | -0.009 |
| 30 | 13.46 | 29 | 0.9938 | 0.037 | -0.010 | 0.034 | 0.001 | 0.080 | 0.013 |
| 36 | 17.07 | 35 | 0.9953 | -0.026 | -0.050 | -0.008 | -0.073 | 0.064 | -0.050 |

Hospital Visits

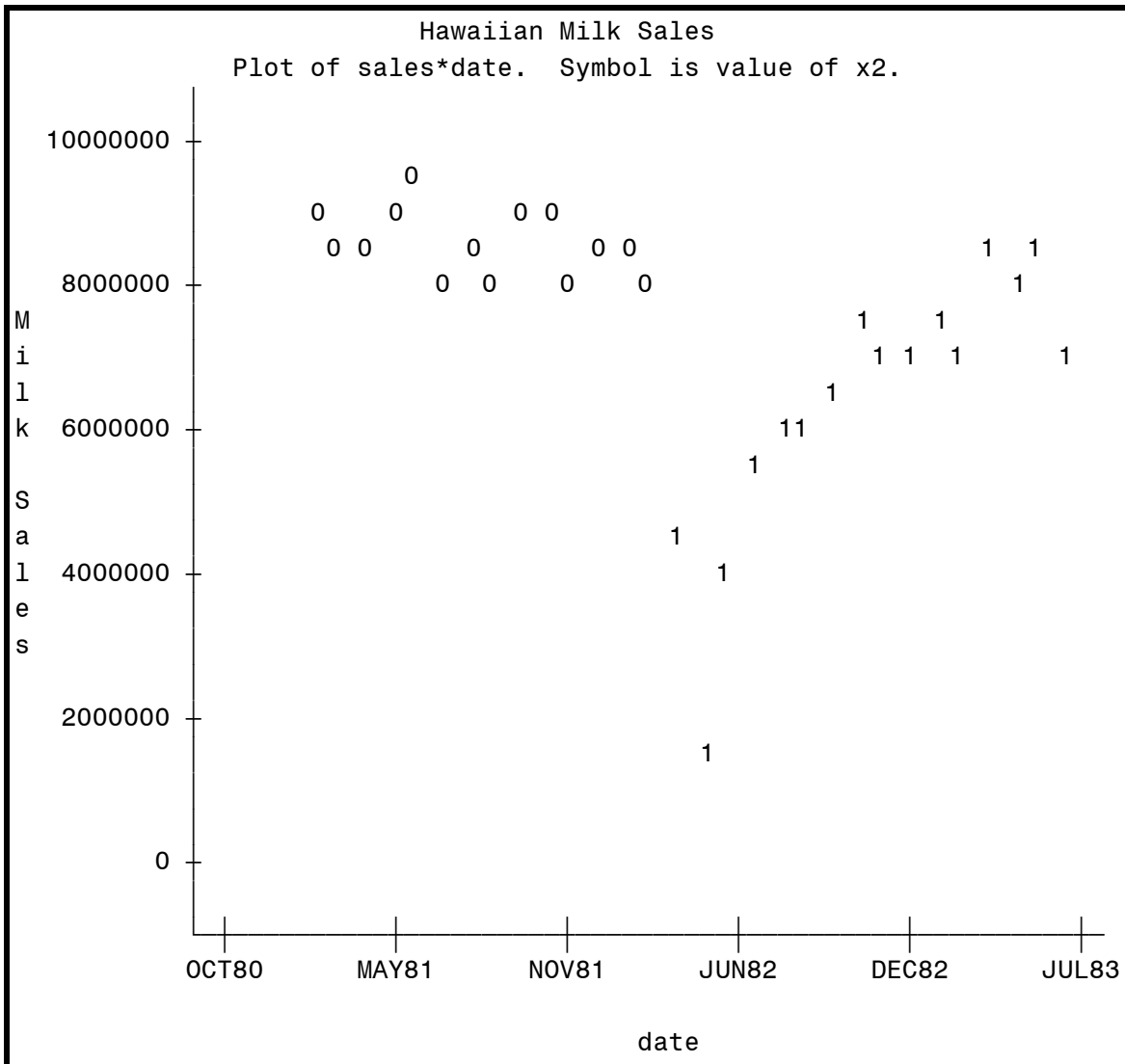


Example: Liu (1988 *Risk Analysis* pg. 689-699) Shows milk sales on Oahu (Hawaii). Tainted (with pesticides) milk was found late in March of 1982. Sales dropped precipitously for the next two months and began to recover.



Graph of the data using the step intervention as a plot symbol:

Our model contains both a point and step intervention. A close-up near the intervention shows that the first big drop-off is in April 1982.



For the point intervention x , we include a denominator (AR analogy) lag to model the exponential approach to the new level. $e \text{ input} = (\quad / (1) x \quad)$

Since the exponential response does not start immediately we add a numerator (MA analogy) lag to our model $e \text{ input} = ((1) / (1) x \quad)$

We also add a shift, $x2$, without a lag operator to see if there is a permanent effect. $e \text{ input} = ((1) / (1) x x2)$

There seems to be some seasonal autocorrelation. We model it with a seasonal multiplicative autoregressive structure $(1 - \alpha_1 B)(1 - \alpha_{12} B^{12})$ and find a little correlation still at lag 2 so we add a moving average term there: $p = (1) (12) \quad q = (2) .$

Here is the resulting code:

```

data liu; input ...
x = (date='01apr82'd); x2 = (date>'01mar82'd);
...
cards;
    FEB77    9331400
(more data)
    JUL83    6906300
;
* Fit intervention model. MA (i.e. numerator) term allows 2
successive dropoffs before recovery ;

proc arima; identify var=sales noprint crosscor=(x x2);
e input=( 1) /(1) x x2 p=(1)(12) q=(2) m1;
run;

```

and the output:

| The ARIMA Procedure | | | | | | |
|-------------------------------|------------|----------------|---------|----------------|--------------|-------|
| Maximum Likelihood Estimation | | | | | | |
| Parameter | Estimate | Standard Error | t Value | Approx Pr > t | Lag Variable | Shift |
| MU | 8674594.0 | 237120.9 | 36.58 | <.0001 | 0 sales | 0 |
| MA1,1 | -0.31264 | 0.12032 | -2.60 | 0.0094 | 2 sales | 0 |
| AR1,1 | 0.29446 | 0.11634 | 2.53 | 0.0114 | 1 sales | 0 |
| AR2,1 | 0.77635 | 0.07040 | 11.03 | <.0001 | 12 sales | 0 |
| NUM1 | -3167440.7 | 324335.4 | -9.77 | <.0001 | 0 x | 0 |
| NUM1,1 | 4855973.9 | 299209.9 | 16.23 | <.0001 | 1 x | 0 |
| DEN1,1 | 0.56562 | 0.03352 | 16.87 | <.0001 | 1 x | 0 |
| NUM2 | -1079129.3 | 222705.0 | -4.85 | <.0001 | 0 x2 | 0 |

| To | Chi-Square | DF | Pr > ChiSq | -----Autocorrelations----- | | | | | | |
|-----|------------|----|------------|----------------------------|--------|--------|--------|--------|--------|--|
| Lag | | | | | | | | | | |
| 6 | 2.46 | 3 | 0.4831 | 0.020 | -0.010 | -0.091 | -0.011 | -0.090 | 0.110 | |
| 12 | 9.47 | 9 | 0.3947 | 0.094 | 0.125 | 0.099 | -0.033 | 0.188 | -0.083 | |
| 18 | 12.27 | 15 | 0.6581 | -0.010 | 0.067 | -0.069 | -0.030 | 0.114 | -0.068 | |
| 24 | 18.29 | 21 | 0.6309 | 0.080 | 0.029 | -0.012 | -0.048 | -0.072 | -0.196 | |

Model for variable sales
 Estimated Intercept 8674594

Autoregressive Factors
 Factor 1: 1 - 0.29446 B**(1)
 Factor 2: 1 - 0.77635 B**(12)

Moving Average Factors
 Factor 1: 1 + 0.31264 B**(2)

Input Number 1
 Input Variable x

Numerator Factors
 Factor 1: -3.17E6 - 4855974 B**(1)

Denominator Factors
 Factor 1: 1 - 0.56562 B**(1)

Permanent Drop: -1079129.3

Additional Temporary Drop: $(-3167440.7 - 4855973.9 B) / (1 - 0.56562 B) X_t$
 where X_t is 1 in April 1982 and 0 elsewhere.

$$\text{Effect}_t = 0.56562 \text{Effect}_{t-1} - 3167441 X_t - 4855974 X_{t-1}$$

Effect/1000

| | | | | | | | | | | | |
|-------------|---|---|-----|-----|-----|-----|-----|----|----|----|-----|
| X | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... |
| Effect/1000 | 0 | 0 | -32 | -66 | -38 | -21 | -12 | -6 | -4 | -2 | |

Demos: Neuse.sas, milk.sas

Effect of Negative Publicity (Contamination) on Milk Sales

