Subset Autoregression: A New Approach

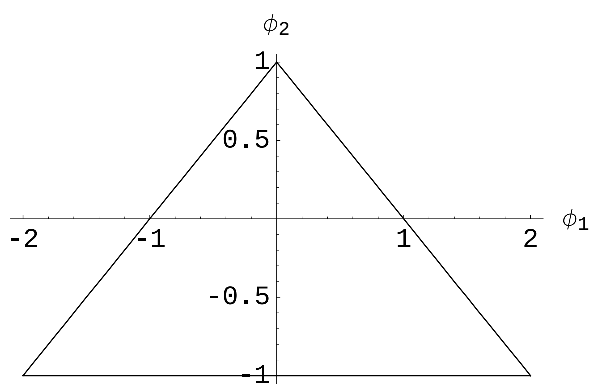
A.Ian McLeod, University of Western Ontario

Ying Zhang, Acadia University

A new family of subset autoregressive models are introduced and a comprehensive approach to model identification, estimation and diagnostic checking is developed for these models. Also a A new version of the partial autocorrelation plot is introduced. These new models are better suited to efficient model building of high-order autoregressions with long time series. Several illustrative examples are given. An R package implementation is available. In many cases subset AR models provide a useful alternative to ARMA models.

AR(p) Model Admissible Region

$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t$, $a_t \sim \mathrm{NID}(0, \sigma_a^2)$ or $\phi(B) x_t = a_t$, B is the backshift operator on t and $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\Phi_p = \{(\phi_1, \ldots, \phi_p) \in \mathbb{R}^p \mid \phi(z) \neq 0,\ z \in \mathbb{C},\ |z| \leq 1\}$. AR(2) region:



■ **AR(3) Admissible Region**

Best Linear Predictor

Given an autocovariance function γ_k, $k = 0, 1, 2, \ldots$ it may easily be shown using calculus that the linear predictor $\phi_{k,1} Z_{t-1} + \ldots + \phi_{k,k} Z_{t-k}$ which minimizes the variance of the error in predicting Z_t is given by the solution to the Yule-Walker equations,

$$\Gamma_p \begin{pmatrix} \phi_{k,1} \\ . \\ . \\ . \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ . \\ . \\ . \\ \gamma_k \end{pmatrix} \qquad (1)$$

where $\Gamma_p = (\gamma_{i-j})_{p\times p}$ is the covariance matrix of p successive time series values. The Durbin-Levinson is a computationally efficient and stable method of solving these special linear equations for $\phi_{k,1}, \ldots, \phi_{k,k}$.

Durbin-Levinson Algorithm

Set $\phi_{1,1} = \gamma_1 / \gamma_0$ and $v_1 = (1 - \phi_{1,1}^2)\, \gamma_0$, where v_k denotes the variance of the k step linear predictor. Then for $k = 2, 3, \ldots$ we can iteratively obtain,

$$\phi_{k,k} = (\gamma_k - \phi_{k-1,1}\, \gamma_{k-1} - \ldots - \phi_{k-1,k-1}\, \gamma_1) / v_{k-1} \qquad \text{DL-1} \qquad (2)$$

$$\begin{pmatrix} \phi_{k,1} \\ . \\ . \\ . \\ \phi_{k,k-1} \end{pmatrix} = \begin{pmatrix} \phi_{k-1,1} \\ . \\ . \\ . \\ \phi_{k-1,k-1} \end{pmatrix} - \phi_{k,k} \begin{pmatrix} \phi_{k-1,k-1} \\ . \\ . \\ . \\ \phi_{k-1,1} \end{pmatrix} \qquad \text{DL-2} \qquad (3)$$

and

$$v_k = v_{k-1}\, (1 - \phi_{k,k}^2). \qquad \text{DL-3} \qquad (4)$$

Other Applications of DL Algorithm

-parameter estimation, Yule-Walker estimates

-parameter estimation via the Burg algorithm

-test for invertibility

-reparameterization of ARMA models for exact MLE

-exact likelihood computation

-exact simulation

-Trench algorithm for efficient computation of Γ_n^{-1}

6 of 41

Reparameterization

Consider an AR(p) with parameters $\phi = (\phi_1, \ldots, \phi_p)$ and let $\zeta = (\zeta_1, \ldots, \zeta_p)$, $\zeta_k = \phi_{k,k}$ where $\phi_{k,k}$, $k = 1, \ldots, p$ are the partial autocorrelations. Barndorff-Nielsen and Schou (1973) showed that $\zeta \longleftrightarrow \phi$ is a bijection which is continuous and differentiable. Hence ζ can be regarded as a reparameterization of the AR in terms of ζ. Efficient algorithms to compute the bijection $\zeta \longleftrightarrow \phi$ are based on the DL recursion.

7 of 41

ARToPacf $\phi \longrightarrow \zeta$

$$\phi_{k-1,i} = (\phi_{k,i} + \phi_{k,k}\,\phi_{k,k-i}) \Big/ \left(1 - \phi_{k,k}^2\right), \quad k = p-1, \ldots, 1; \; i = 1, \ldots, k-1 \qquad (5)$$

Derivation

From DL recursion we can write,

$$\phi_{k-1,i} = \phi_{k,i} + \phi_{k,k}\,\phi_{k-1,k-i} \qquad (6)$$

by symmetry,

$$\phi_{k-1,\,k-i} = \phi_{k,\,k-i} + \phi_{k,k}\,\phi_{k-1,\,i} \tag{7}$$

Subing (8) in (7) and simplifying yields (6).

8 of 41

PacfToAR $\zeta \longrightarrow \phi$

$$\phi_{k,\,i} = \phi_{k-1,\,i} - \zeta_k\,\phi_{k-1,\,k-i}; \quad k = 2, \ \ldots, \ p; \ i = 1, \ \ldots, \ k-1. \tag{8}$$

This follows directly from eqn. (4).

9 of 41

Subset AR Models:

Principle of Parameter Parsimony suggests considering models such as,

$x_t = \phi_1\,x_{t-1} + \phi_2\,x_{t-2} + \phi_9\,x_{t-9} + a_t$

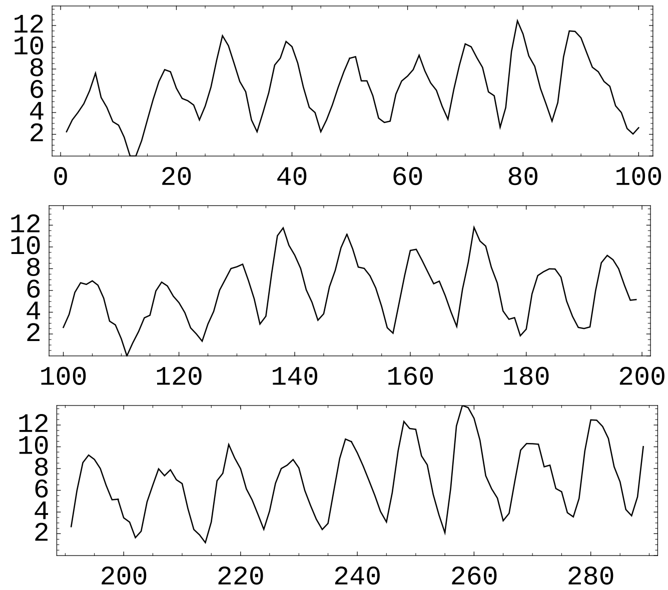
These models may be fit using least-squares. Least squares subset regression algorithms may be used. In general the $\mathrm{AR}_\phi(i_1,\,i_2,\,\cdots,\,i_m)$ is defined by

$$x_t = \phi_{i_1}\,x_{t-i_1} + \phi_{i_2}\,x_{t-i_2} + \cdots + \phi_{i_m}\,x_{t-i_m} + a_t \tag{9}$$

10 of 41

Annual Sunspot Series, 1700-1988

Consider a power transformation z^λ for $\lambda = 1,\ 0.5,\ 0.33$; $g_3 = 1.02,\ 0.18,\ -0.25$ so a square-root transformation is selected.



11 of 41

AR(2) - Exact MLE

	Parameter	SE	Z
ϕ_1	1.38858	0.0425451	32.6378
ϕ_2	−0.690569	0.0425451	−16.2314
μ	48.6135	3.22231	15.0866
σ_a	16.5429		

{PortmanteauStatistic→60.4823,MaxLag→25,Pvalue→0.0000325296}

12 of 41

AR$_\phi$(1, 2, 9) - LS

We fit AR$_\phi$(1, 2, i) for i = 8, 9, 10, 11, 12. Only with i = 9 was an acceptable fit obtained.

	Parameter	SE	Z
ϕ_1	1.24378	0.0588235	21.1442
ϕ_2	−0.523923	0.0938781	−5.58089
ϕ_9	0.201266	0.0588235	3.42153
μ	6.34343	0.794134	7.98786
σ_a	1.06488		

```
{PortmanteauStatistic→28.4572, MaxLag→25, Pvalue→0.16102}
```

13 of 41

AR(1, 2, 4)

Is the transformation $\phi \longleftrightarrow \zeta$ useful for exact MLE of subset models?

We were able to extend the Theorem of Barndorff-Nielsen and Schou (1973) to show that the transformation $(\phi_{i_1}, \cdots, \phi_{i_m}) \longrightarrow (\zeta_{i_1}, \cdots, \zeta_{i_m})$ is a one-to-one and is continuous and differentiable. But it is very complicated and not possible to compute easily. Also the transformation is not onto.

■ Admissible Region of $AR_\phi(1, 4)$ in ζ space

■ Admissible Region of $AR_\phi(1, 2, 4)$ in ζ space

14 of 41

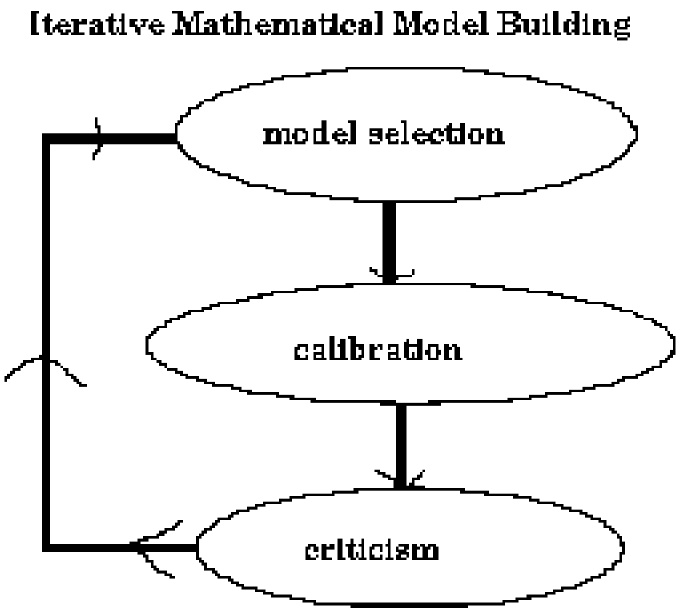
New Subset Models

$(\zeta_{i_1}, \cdots, \zeta_{i_m}) \longrightarrow (\phi_1, \cdots, \phi_p)$, $i_m = p$, defined using `PacfToAR` given in eqn. (9) and letting $\zeta_i \in (-1, 1)$. The admissible region is simply a cube in *m*-dimensions.

And we can use the `PacfToAR` and `ARToPacf` transformations.

15 of 41

Model Building



16 of 41

Exact Loglikelihood

$$L(\phi, \sigma_a^2) = -\frac{1}{2} \operatorname{Log}(\det(\Gamma_n)) - \frac{1}{2} x' \Gamma_n^{-1} x \tag{10}$$

where Γ_n is the covariance matrix of $x = (x_1, \ldots, x_n)$ and $\phi = (\phi_1, \ldots, \phi_p)$.

Champernowne (1948) showed that

$$x' \Gamma_n^{-1} x = \beta' D \beta / \sigma_a^2, \tag{11}$$

where $\beta = (-1, \phi_1, \ldots, \phi_p)$ and D is the $(p+1)$-by-$(p+1)$ matrix with (i, j)-entry

$$D_{i,j} = D_{j,i} = x_i x_j + \ldots + x_{n+1-i} x_{n+1-j}. \tag{12}$$

17 of 41

Exact Loglikelihood (con't)

The standardized covariance determinant of order p, $g_p = \det(\Gamma_p / \sigma_a^2)$ where $\Gamma_p = (\gamma_{i-j})$ and $\gamma_k = \mathrm{Cov}(x_t, x_{t-k})$ may be written (Barndorff-Nielsen and Schou, 1973, eqns. 5, 8) as

$$g_p = \prod_{j=1}^{p} (1 - \zeta_j^2)^{-j}. \tag{13}$$

Hence the exact loglikelihood function (5.3.1) may now be written,

$$L(\phi, \sigma_a^2) = -\frac{n}{2} \mathrm{Log}(\sigma_a^2) - \frac{1}{2} \mathrm{Log}(g_p) - \frac{1}{2\sigma_a^2} S(\phi) \tag{14}$$

where $S(\phi) = \beta' D \beta$.

Concentrated Loglikelihood

Maximizing (5.3.1), $L(\phi, \sigma_a^2)$, over σ_a^2 , we obtain

$$\hat{\sigma}_a^2 = S(\phi) / n \tag{15}$$

and the profile loglikelihood for ϕ can be written,

$$L(\phi) = -\frac{n}{2} \mathrm{Log}(S(\phi) / n) - \frac{1}{2} \mathrm{Log}(g_p). \tag{16}$$

After the initial computation of the matrix D which only needs to be done once, each further evaluation of the likelihood $L(\phi)$ only requires $O(p^2)$ flops. Provided that $p << n$, this is much faster than other exact likelihood algorithms for ARMA

Reparameterized Concentrated Loglikelihood

To maximize $L(\phi)$, it is convenient to re-parameterize using the $\zeta = (\zeta_1, \ldots, \zeta_p)$ parameters. We can then write,

$$L(\phi(\zeta)) = -\frac{n}{2}\,\mathrm{Log}(S(\phi(\zeta))/n) - \frac{1}{2}\,\mathrm{Log}(g_p). \tag{17}$$

20 of 41

The Burg Estimates

-first order efficient

-fast to compute using DL algorithm

-always in admissible region

Percival and Walden (1993, §9.5) give a new statistical derivation of the Burg algorithm.

Zhang & McLeod (2005) showed that the Burg estimates have that the first-order bias of the Burg estimates is the same as least-squares for AR(p), $p = 1, 2, 3$.

21 of 41

Derivation of the Burg Estimates

In this algorithm $\vec{e}_t(k)$ and $\overleftarrow{e}_t(k)$ denote the forward and backward k-th order linear prediction errors for z_t based on $z_{t-1}, \ldots, z_{t-k}$ and $z_{t+1}, \ldots, z_{t+k}$ respectively. The Burg algorithm is characterized by the fact that $\hat{\phi}_{k,k}$ minimizes

$$\mathrm{SS}_k = \sum_{t=k+1}^{n} \vec{e}_t^{\,2}(k-1) + \overleftarrow{e}_{t-k}^{\,2}(k-1)\,. \tag{18}$$

It may be shown that the $\hat{\phi}_{k,k}$ which minimize SS_k is given by,

$$\hat{\phi}_{k,k} = A_k / B_k \tag{19}$$

$$A_k = 2\,\{\textstyle\sum_{t=k+1}^{n} \vec{e}_t(k-1)\,\overleftarrow{e}_{t-k}(k-1)\} \tag{20}$$

$$B_k = \sum_{t=k+1}^{n} \left\{\vec{e}_t^{\,2}(k-1) + \overleftarrow{e}_{t-k}^{\,2}(k-1)\right\} \tag{21}$$

22 of 41

The Burg Algorithm

The Burg algorithm produces the partial autocorrelation estimates directly from the data. We assume z_t has mean zero or has been mean corrected. Then to fit an AR(p),

Step 1: Select p. Initialization. Set $k = 1$ and for $t = 2, \ldots, n$,

$$\vec{e}_t(k-1) = z_t, \tag{22}$$

$$\overleftarrow{e}_{t-1}(k-1) = z_{t-1}, \tag{23}$$

Step 2: Compute $\hat{\phi}_{k,k}$ using eqn. (11).

Step 3: Update. For $t = k+1, \ldots, n$,

$$\vec{e}_t(k) = \vec{e}_t(k-1) - \hat{\phi}_{k,k}\, \overleftarrow{e}_{t-k}(k-1), \tag{24}$$

$$\overleftarrow{e}_{t-k}(k) = \overleftarrow{e}_{t-k}(k-1) - \hat{\phi}_{k,k}\, \vec{e}_t(k-1). \tag{25}$$

Step 4: If $k = p$, terminate otherwise set $k = k+1$. Repeat Steps 2-4.

23 of 41

Information Matrix for AR_ζ

Let I_ϕ and I_ζ denote the large-sample information matrix per observation for the parameters ϕ and ζ respectively in an AR(p). Then

$$I_\zeta = \mathbb{J}' \, I_{\phi(\zeta)} \, \mathbb{J}, \tag{26}$$

$$\mathbb{J} = \frac{\partial \phi}{\partial \zeta} = \prod_{k=1}^{p-1} \mathbb{J}_{p-k} \tag{27}$$

where $\mathbb{J}_{p-k}$ is derived below. In the subset case, the corresponding rows and columns of I_ζ are selected.

24 of 41

Derivation of $\mathbb{J}_{p-k}$

The required Jacobian may be derived from the sequence of transformations, starting with $\phi_i = \phi_{p,i}$, $i = 1, \ldots p$ and continuing until we reach $\zeta_i = \phi_{i,i}$, $i = 1, \ldots, p$:

$$\mathbb{T}_{p-1}: \{\phi_{p,1}, \phi_{p,2}, \ldots, \phi_{p,p-1}, \phi_{p,p}\} \longleftrightarrow \{\phi_{p-1,1}, \phi_{p-1,2}, \ldots, \phi_{p-1,p-1}, \phi_{p,p}\} \tag{28}$$

$$\mathbb{T}_{p-2}: \{\phi_{p-1,1}, \phi_{p-1,2}, \ldots, \phi_{p-1,p-1}, \phi_{p,p}\} \longleftrightarrow \{\phi_{p-2,1}, \phi_{p-2,2}, \ldots, \phi_{p-2,p-3}, \phi_{p-1,p-1}, \phi_{p,p}\} \tag{29}$$

....

$$\mathbb{T}_1: \{\phi_{2,1}, \phi_{2,2}, \ldots, \phi_{p-1,p-1}, \phi_{p,p}\} \longleftrightarrow \{\phi_{1,1}, \phi_{2,2}, \ldots, \phi_{p-1,p-1}, \phi_{p,p}\} \tag{30}$$

The general form of these transformations is given by the Durbin-Levinson recursion,

$$\phi_{p,j} = \phi_{p-1,j} - \phi_{p-1,p-j}\,\phi_{p,p}, \quad j = 1, \ldots, p-1$$

Derivation of $\mathbb{J}_{p-k}$ (con't)

In general, the Jacobian of the transformation, $\mathbb{T}_{p-k}$, may be written as a partitioned matrix,

$$\mathbb{J}_{p-k} = \begin{pmatrix} J_{p-k} & A_{p-k,k} \\ 0_{p-k} & I_k \end{pmatrix} \tag{31}$$

$0_{p-k} = (0)_{k,\, p-k}$, I_k is the $k \times k$ identity matrix and $A_{p-k,k}$ is a $(p-k) \times k$ matrix whose first column is

$$\{-\phi_{p-k,p-k}, -\phi_{p-k,p-k-1}, \ldots, -\phi_{p-k,1}\}$$

and whose remaining elements are all 0. The matrix J_{p-k} may be written explicitly as the $(p-k) \times (p-k)$ matrix with (i, j)-entry $a(p, k)$ where,

$$a(p, k) = \begin{cases} 1 & \text{if } i = j \\ -\zeta_{p-k+1} & \text{if } i = p-k+1-j \wedge i \neq j \\ 1-\zeta_{p-k+1} & \text{if } i = p-k+1-j \wedge i = j \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

Example

When $p = 4$,

$$\mathbb{J}_3 = \begin{pmatrix} 1 & 0 & -\zeta_4 & -\zeta_3 \\ 0 & 1-\zeta_4 & 0 & -\zeta_2 - \zeta_1 (1+\zeta_2) \zeta_3 \\ -\zeta_4 & 0 & 1 & -\zeta_1 (1+\zeta_2) - \zeta_2 \zeta_3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbb{J}_2 = \begin{pmatrix} 1 & -\zeta_3 & -\zeta_2 & 0 \\ -\zeta_3 & 1 & -\zeta_1 (1+\zeta_2) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbb{J}_1 = \begin{pmatrix} 1-\zeta_2 & -\zeta_1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Simulation Experiment

$\zeta = \{0.5,\ 0.5,\ 0.5,\ 0.5\}$ and simulated 1,000 realizations of a time series with length $n = 1000$.

The observed sample covariance matrix of $\hat{\zeta}$ in the simulations was,

$$\begin{pmatrix} 0.0133474 & 0.00452826 & 0.00180141 & 0.000421101 \\ 0.00452826 & 0.00250702 & 0.000142564 & 0.000209351 \\ 0.00180141 & 0.000142564 & 0.000794065 & 0.0000592178 \\ 0.000421101 & 0.000209351 & 0.0000592178 & 0.000758274 \end{pmatrix}$$

and the theoretical large-sample approximation given by $I_{\zeta}^{-1} / 1000$

$$\begin{pmatrix} 0.01425 & 0.0045 & 0.0015 & 0 \\ 0.0045 & 0.00225 & 0. & 0 \\ 0.0015 & 0. & 0.00075 & 0 \\ 0 & 0 & 0 & 0.00075 \end{pmatrix}.$$

Large-Sample Distribution of MLE

Theorem 1

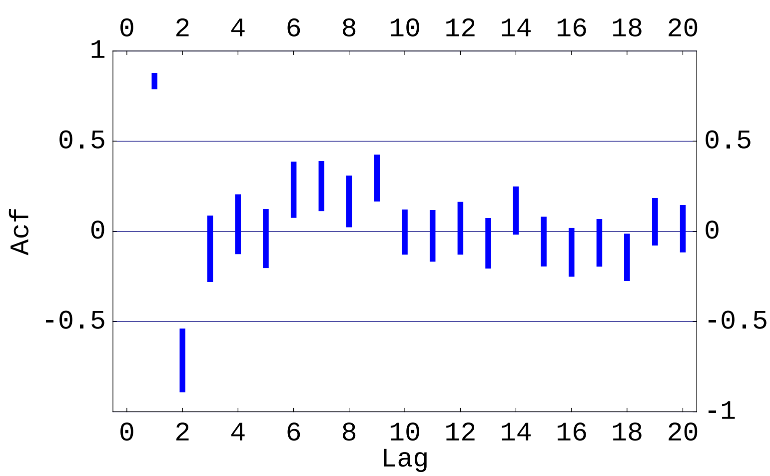
$\hat{\zeta} \xrightarrow{p} \zeta$ and $\sqrt{n}\,(\hat{\zeta} - \zeta) \xrightarrow{L} N(0, I_{\zeta}^{-1})$

■ **Proof**

29 of 41

Partial Autocorrelation Plot

Using the Burg or exact MLE we obtain $\hat{\zeta}_k = \hat{\phi}_{k,k}$ and then using Theorem 1 obtain EstSd($\hat{\zeta}_k$). Plot the 95% intervals $\hat{\zeta}_k \pm 1.96\,\text{EstSd}(\hat{\zeta}_k)$.



30 of 41

AIC/BIC Model Selection

For $\text{AR}_\zeta(i_1, \cdots, i_m)$, $\mathcal{L} = (-n/2)\log(\hat{\sigma}_a^2)$

Since $\hat{\sigma}_a^2 \approx c_0 \left(1 - \hat{\zeta}_{i_1}^2\right) \cdots \left(1 - \hat{\zeta}_{i_m}^2\right)$

$$\text{BIC}(i_1, \cdots, i_m) = n \log \prod_{k \in \{i_1, \cdots, i_m\}} \left(1 - \hat{\zeta}_k^2\right) + m \log(n) \tag{33}$$

We don't need to search all subsets. Just arrange $\hat{\zeta}_k^2$ in ascending order and proceed with the evaluation of the BIC.

Similarly for other IC.

Annual Sunspot Example

k	AIC	BIC	Model
1	−330.899	−327.232	{1}
2	−532.767	−525.435	{1, 2}
3	−548.387	−537.388	{1, 2, 8}
4	−561.591	−546.926	{1, 2, 8, 7}
5	−573.957	−555.625	{1, 2, 8, 7, 9}
6	−582.986	−560.987	{1, 2, 8, 7, 9, 6}
7	−588.369	**−562.704**	{1, 2, 8, 7, 9, 6, 17}
8	−590.1	−560.769	{1, 2, 8, 7, 9, 6, 17, 3}
9	**−590.659**	−557.661	{1, 2, 8, 7, 9, 6, 17, 3, 15}
10	−590.281	−553.617	{1, 2, 8, 7, 9, 6, 17, 3, 15, 18}
11	−589.845	−549.515	{1, 2, 8, 7, 9, 6, 17, 3, 15, 18, 14}
12	−589.358	−545.361	{1, 2, 8, 7, 9, 6, 17, 3, 15, 18, 14, 16}

MinAICModel → {1, 2, 8, 7, 9, 6, 17, 3, 15}

MinBICModel → {1, 2, 8, 7, 9, 6, 17}

Fitted AR_ζ for Annual Sunspots

	Parameter	SE	Z
ζ_1	0.839169	0.281783	2.97807
ζ_2	-0.671837	0.0824506	-8.14835
ζ_6	0.252221	0.0620332	4.0659
ζ_7	0.231398	0.0608168	3.80484
ζ_8	0.196152	0.0629042	3.11826
ζ_9	0.300116	0.0703228	4.26769
ζ_{17}	-0.0730557	0.0604646	-1.20824
μ	6.34343	0.657045	9.65448
σ_a	1.04209		

33 of 41

Distribution Residual Autocorrelations

After fitting the residuals $\hat{a}_t$ are used to check the important assumption of independence. $\hat{r} = (r_{\hat{a}}(1), \cdots, r_{\hat{a}}(L))$

Theorem 2

$$\sqrt{n}\,\hat{r} \xrightarrow{L} N(0, \mathcal{V}), \mathcal{V} = I_m - \mathcal{X} J_\zeta I_\zeta^{-1} J_\zeta' \mathcal{X}'$$

where $\mathcal{X}$ is the $L \times m$ matrix with (i, j)-entry ψ_{i-j}, where ψ_k is the coefficient of B^k in the expansion $1/\phi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$ and $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$.

Remark 1: Since $J_\zeta' \mathcal{X}' \mathcal{X} J_\zeta \simeq J_\zeta' I_\zeta^{-1} J_\zeta$, $\mathcal{V}$ is approximately idemptotent with rank $L - m$.

Remark 2: In the case of squared residuals, the autocorrelations are NID$(0, 1/n)$.

34 of 41

Ljung-Box Test

Since $\mathcal{V}$ is approximately idemptotent with rank $L - m$ and hence we can use the Ljung-Box test

$$Q_L = n(n+2) \sum_{k=1}^{L} \hat{r}(k)^2 \Big/ (n-k) \qquad (34)$$

Under the null hypothesis of model adquacy, Q_L is approximately χ^2-distributed with $L - m$ df.

35 of 41

Sunspot Series Example

Ljung-Box Test

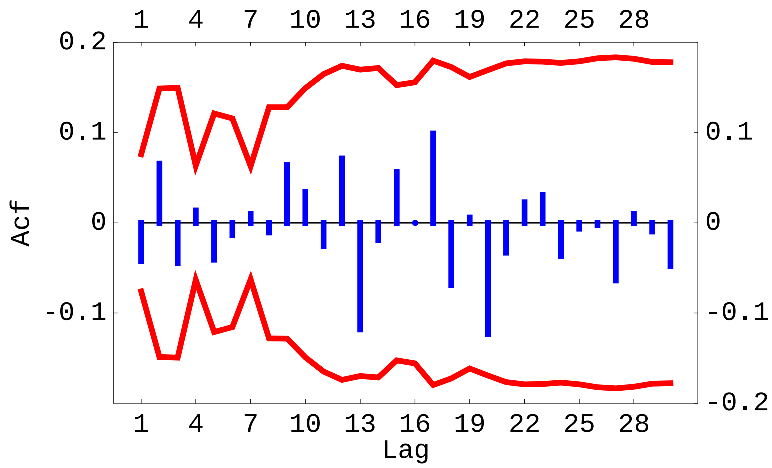
Q_L	L	p-value
21.	20	0.05
22.3	25	0.18
24.4	30	0.33
25.5	35	0.55
30.3	40	0.55
32.	45	0.7
33.3	50	0.83

36 of 41

Residual Autocorrelation Plot

As noted by Hosking and Ravishanker (1993) the Bonferonni Inequality may be used to obtain 5% simultaneous significance levels.

■ **AR$_\zeta$(1, 2, 6, 7, 8, 9, 17)**



37 of 41

Annual Sunspot Series

■ **Squared Residuals Ljung-Box Test**

Q_L	L	p-value
35.2	20.	0.00043
40.5	25.	0.00111
47.2	30.	0.00136
53.8	35.	0.00161
63.4	40.	0.00078
69.5	45.	0.00097
70.8	50.	0.00358

Conditional heteroscedastic variation is present. Nonlinear model needed.

38 of 41

Monthly Sunspot Series, 1749-1983, $n = 2820$

Taking $L = 300$ and $M = 100$ the best AIC$_\zeta$ and BIC$_\zeta$ models were determined. The AIC and BIC were also used with the usual AR(p). The fits are summarized:

Model	IC	m	$\mathcal{L}$	AIC	BIC
AR_ζ	AIC	70	−148.2	436.4	852.6
AR_ζ	BIC	20	−236.5	513.0	631.9
AR	AIC	27	−241.1	536.3	696.8
AR	BIC	21	−252.5	547.0	671.8

39 of 41

Appendix: Bonferonni Inequality

First consider white noise case, $r_k \sim \mathrm{NID}(0, 1/n)$. Then

$\Pr\left\{\left|r_k\right| < c \big/ \sqrt{n},\ k = 1, \cdots, L\right\} = 1 - \alpha$

$\Pr\{|Z_k| < c,\ k = 1, \cdots, L\} = 1 - \alpha$

where $Z_k = r_k \big/ \sqrt{n} \sim \mathrm{NID}(0, 1)$. From Basic Result in elementary probability,

$\Pr\{|Z_k| < c,\ k = 1, \cdots, L\} = 1 - \prod_{k=1}^{L} \Pr\{|Z_k| > c\}$

$(1 - 2\,(1 - \Phi(c)))^L = 1 - \alpha$

$\therefore c = \Phi^{-1}((1 + (1 - \alpha)^{1/L})/2)$

L	1	2	10	20	40	60
c	1.96	2.24	2.80	3.02	3.22	3.33

40 of 41

Bonferonni Inequality

$1 \geq \Pr\{\xi_1 \bigcup \xi_2\} = \Pr\{\xi_1\} + \Pr\{\xi_2\} - \Pr\{\xi_1 \bigcap \xi_2\}$

$1 + \Pr\{\xi_1 \bigcap \xi_2\} \geq \Pr\{\xi_1\} + \Pr\{\xi_2\} = (1 - \Pr\{\overline{\xi}_1\}) + (1 - \Pr\{\overline{\xi}_2\})$

$\Pr\{\xi_1 \bigcap \xi_2\} \geq 1 - (\Pr\{\overline{\xi}_1\} + \Pr\{\overline{\xi}_2\})$

-general case established by induction

-higher-order expansion

41 of 41

Bonferonni Approximation

$$\bigcap_{i=1}^{L} \Pr\{\xi_i\} \approx 1 - \sum_{i=1}^{L} \Pr\{\overline{\xi}_i\} \qquad (35)$$

taking $\xi_i = \{ |\hat{r}_i| < c \, \mathrm{EstSd}(\hat{r}_i)\}$

$\bigcap_{i=1}^{L} \Pr\{\overline{\xi}_i\} = 1 - \alpha$

$\Pr\{\overline{\xi}_i\} = \alpha / m$ so $c = \Phi^{-1}(1 - \alpha / (2 L))$

L	1	2	10	20	40	60
c	1.96	2.24	2.80	3.02	3.22	3.34

Almost but not exactly the same!