Modeling with ARMA Processes: I

- goal: determine which ARMA(p, q) process is best model for observed time series $x_1, \ldots, x_n$
- tasks at hand include
  - determining p and q (order selection)
  - estimating process mean (easily done!), coefficients ϕ_j & θ_j (not so easy) and white noise variance σ^2 (relatively easy)
  - subjecting selected model to goodness-of-fit tests
- note: will assume that, if need be, series $x_1, \ldots, x_n$ has been adjusted so that it can be regarded as realization of zero mean stationary process (usual procedure: take sample mean $\bar{x}' = \frac{1}{n}\sum_{t=1}^{n} x'_t$ of original series $x'_1, \ldots, x'_n$ and set $x_t = x'_t - \bar{x}'$)

Modeling with ARMA Processes: II

- with p & q assumed initially to be known, will advocate Gaussian maximum likelihood (ML) estimators for ϕ_j, θ_j & σ^2
- requires use of nonlinear optimization procedure, for which need good initial estimates of coefficients ϕ_j & θ_j
- can base initial estimates on easier-to-compute estimators
  - Yule–Walker (Y–W) estimator (good for AR(p) case)
  - Burg estimator (for AR(p) also)
  - innovations algorithm (handles MA(q) and ARMA(p, q))
  - Hannan–Rissanen (adapts Y–W to handle ARMA(p, q))
- will now describe these estimators and also give some preliminary thoughts about order selection

BD: 121–122, CC: 149, SS: 113

Yule–Walker Estimation: I

- assume causal (and hence stationary) AR(p) model:

$$\phi(B)X_t = Z_t, \text{ i.e., } X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t, \qquad (*)$$

  with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- can develop set of p linear equations linking

$$\boldsymbol{\phi} = [\phi_1, \ldots \phi_p]'$$

  to ACVF values $\gamma(0)$, $\gamma(1)$, ..., $\gamma(p-1)$
- Y–W estimators gotten by substituting usual estimator $\hat{\gamma}(h)$ for $\gamma(h)$ in p equations (so-called 'moment matching' (MM))
- one additional equation needed to estimate σ^2 via this scheme

Yule–Walker Estimation: II

- with $h \geq 0$, multiply both sides of $(*)$ by X_{t-h}:

$$X_t X_{t-h} - \sum_{j=1}^{p} \phi_j X_{t-j} X_{t-h} = Z_t X_{t-h}$$

- take expectations to get

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = E\{Z_t X_{t-h}\} = \begin{cases} \sigma^2, & h = 0; \\ 0, & h \geq 1, \end{cases} \quad (**)$$

because causality allows us to write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ and hence } E\{Z_t X_{t-h}\} = \sum_{j=0}^{\infty} \psi_j E\{Z_t Z_{t-h-j}\}$$

(recall that $\psi_0 = 1$)

Yule–Walker Estimation: III

- leads to $\sum_{j=1}^{p} \phi_j \gamma(h-j) = \gamma(h)$ for $h = 1, \ldots, p$:

$$\begin{aligned} \phi_1\gamma(0) + \phi_2\gamma(1) + \cdots \phi_p\gamma(p-1) &= \gamma(1) \\ \phi_1\gamma(1) + \phi_2\gamma(0) + \cdots \phi_p\gamma(p-2) &= \gamma(2) \\ &\vdots \\ \phi_1\gamma(p-1) + \phi_2\gamma(p-2) + \cdots \phi_p\gamma(0) &= \gamma(p) \end{aligned}$$

- in matrix notation we have $\Gamma_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p$, where

$$\Gamma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix}, \quad \boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}, \quad \boldsymbol{\gamma}_p = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{bmatrix}$$

- to (mis)quote Yogi Berra (1925–2015):

  ‘This is like déjà vu all over again!’

Yule–Walker Estimation: IV

- *exactly* same matrix equation arose when trying to find coefficients of best linear predictor of X_{n+1} given $X_n, \ldots, X_1$ for a general stationary process $\{X_t\}$ (i.e., not necessarily AR(p))
- given time series $x_1, \ldots, x_n$, form usual estimate of ACVF:

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

- with $\hat{\Gamma}_p$ & $\hat{\boldsymbol{\gamma}}_p$ formed by replacing $\gamma(h)$'s in Γ_p & $\boldsymbol{\gamma}_p$ by $\hat{\gamma}(h)$'s, Y–W estimator $\hat{\boldsymbol{\phi}}$ of AR(p) coefficients $\boldsymbol{\phi}$ given by

$$\hat{\boldsymbol{\phi}} = \hat{\Gamma}_p^{-1} \hat{\boldsymbol{\gamma}}_p,$$

where inverse $\hat{\Gamma}_p^{-1}$ exists as long as time series isn't 'boring'

Yule–Walker Estimation: V

- can solve equation using Levinson–Durbin recursions
- note: provides Y–W estimates not only for order p, but also for all lower orders $1, \ldots, p-1$
- once $\hat{\boldsymbol{\phi}}$ has been computed, can return to $h = 0$ case of $(**)$, namely,

$$\sigma^2 = \gamma(0) - \sum_{j=1}^{p} \phi_j \gamma(j), \text{ to get estimator } \hat{\sigma}^2 = \hat{\gamma}(0) - \sum_{j=1}^{p} \hat{\phi}_j \hat{\gamma}(j)$$

(similar equation arose for getting MSE of best linear predictor)

- fitted model

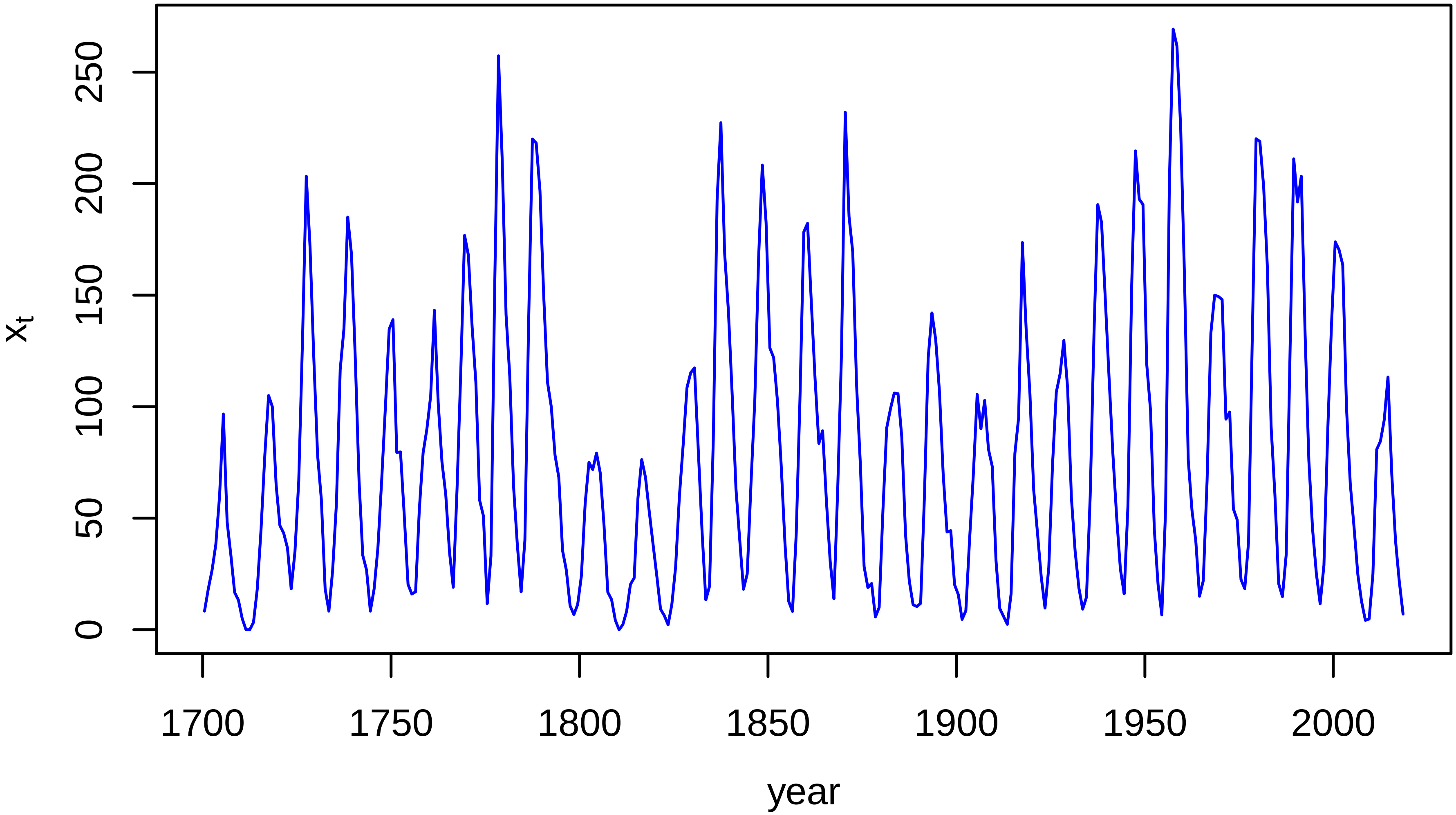
$$X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} = Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \hat{\sigma}^2),$$

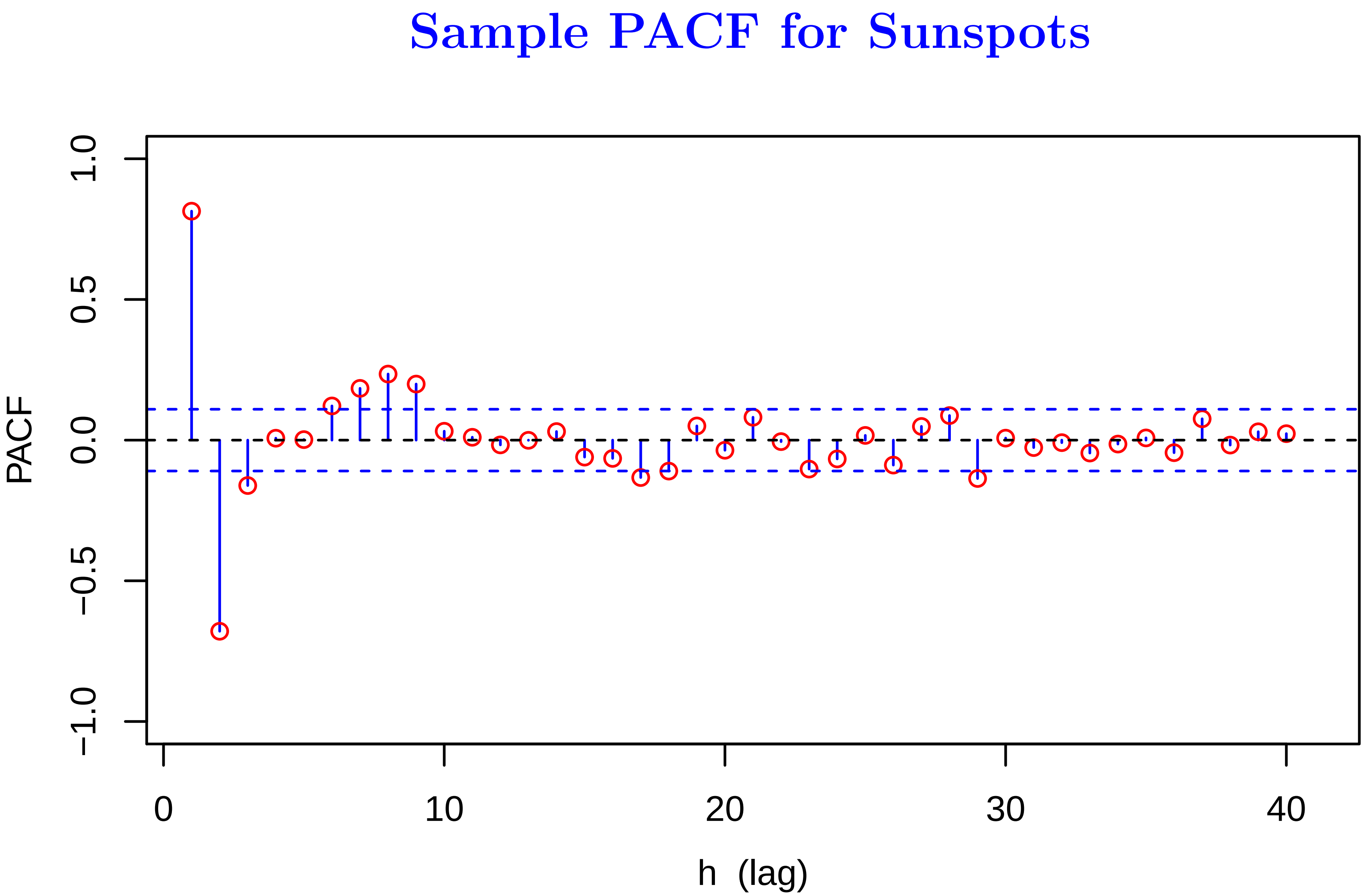
is *guaranteed* to be causal and hence stationary!

Yule–Walker Estimation: VI

- fitted model has theoretical ACVF that is *identical* to estimates $\hat{\gamma}(h)$ at lags $h = 0, 1, \ldots, p$, but in general is different at higher lags
- as an example, let's revisit the sunspot time series
  - use Y–W to fit AR models of orders corresponding to large values in sample PACF: $p = 1$, 2, 3, 6, 7, 8, 9, 17, 18 and 29
  - compare sample ACVF to theoretical ACVFs corresponding to 10 fitted AR models

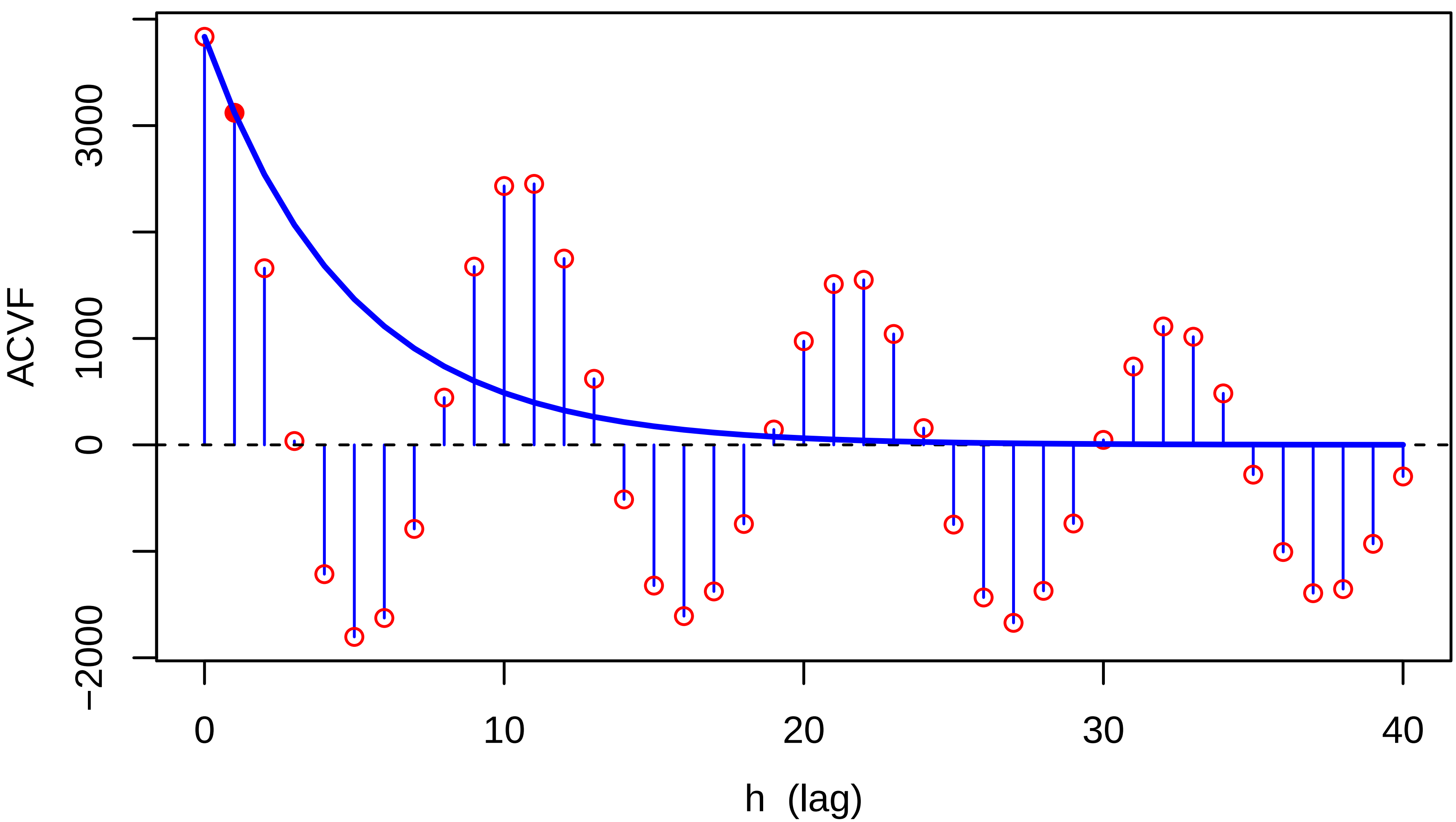
Sunspots (1700–2018)



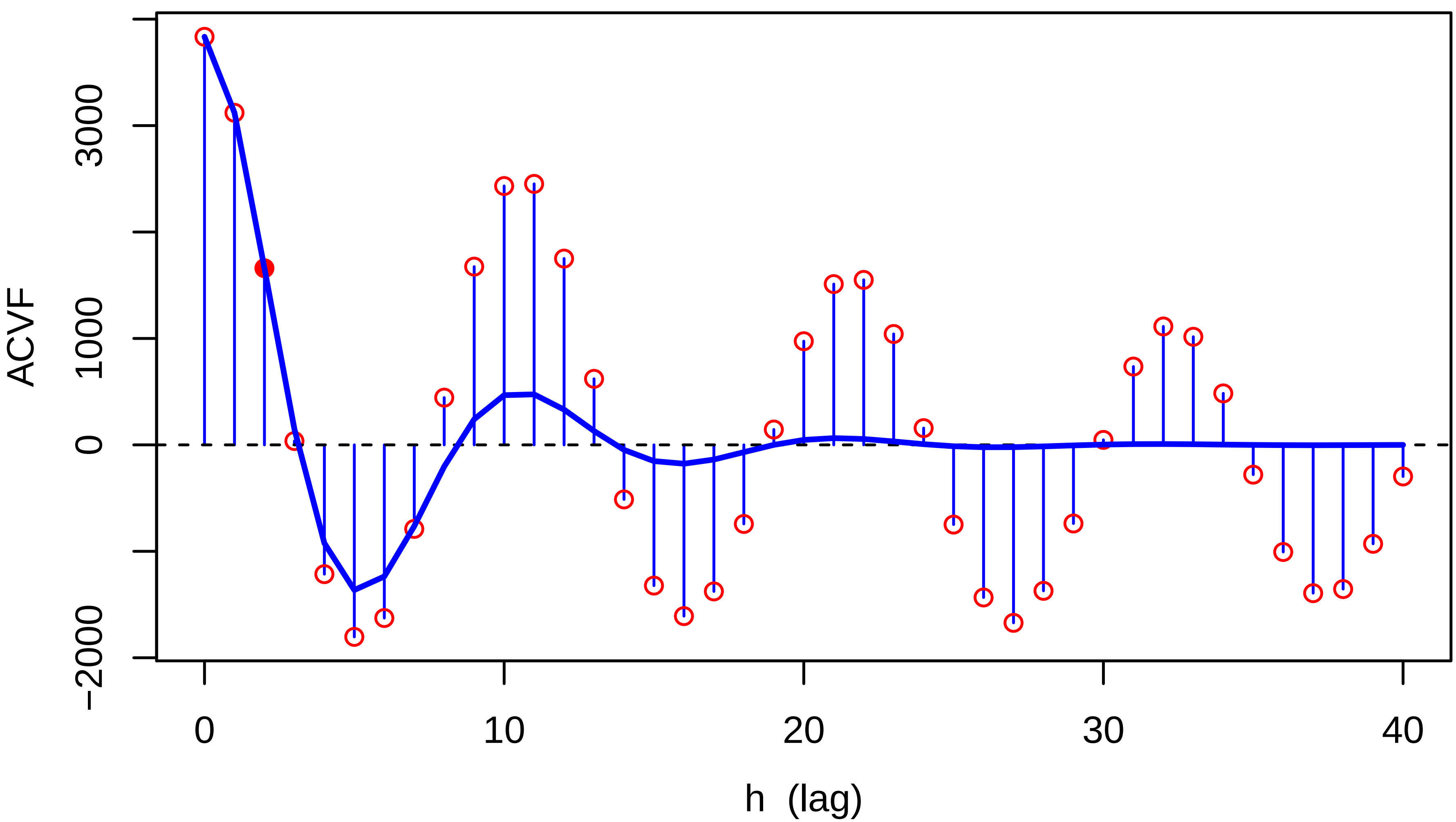


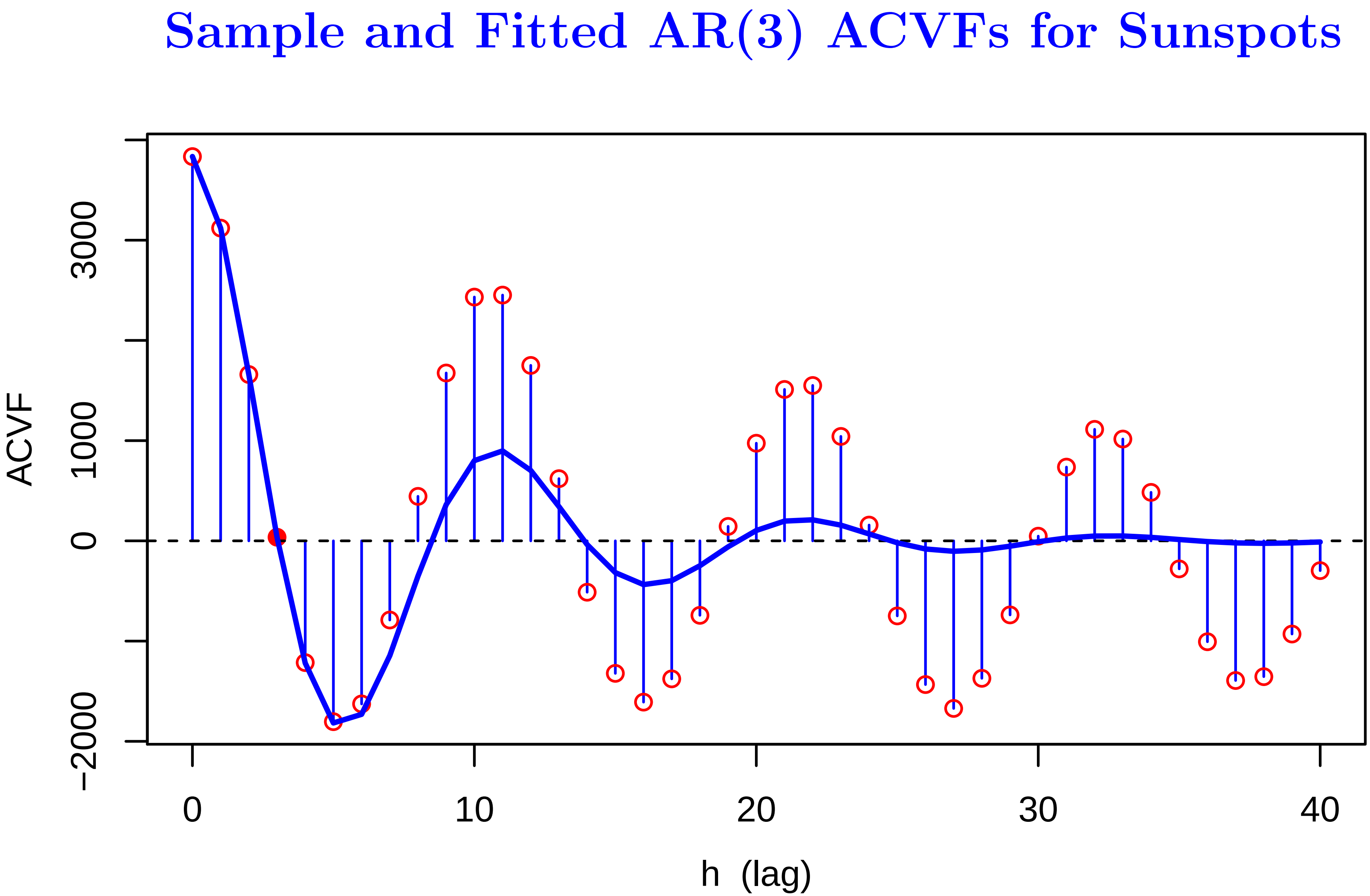
Sample PACF for Sunspots
PACF
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
h (lag)

Sample and Fitted AR(1) ACVFs for Sunspots



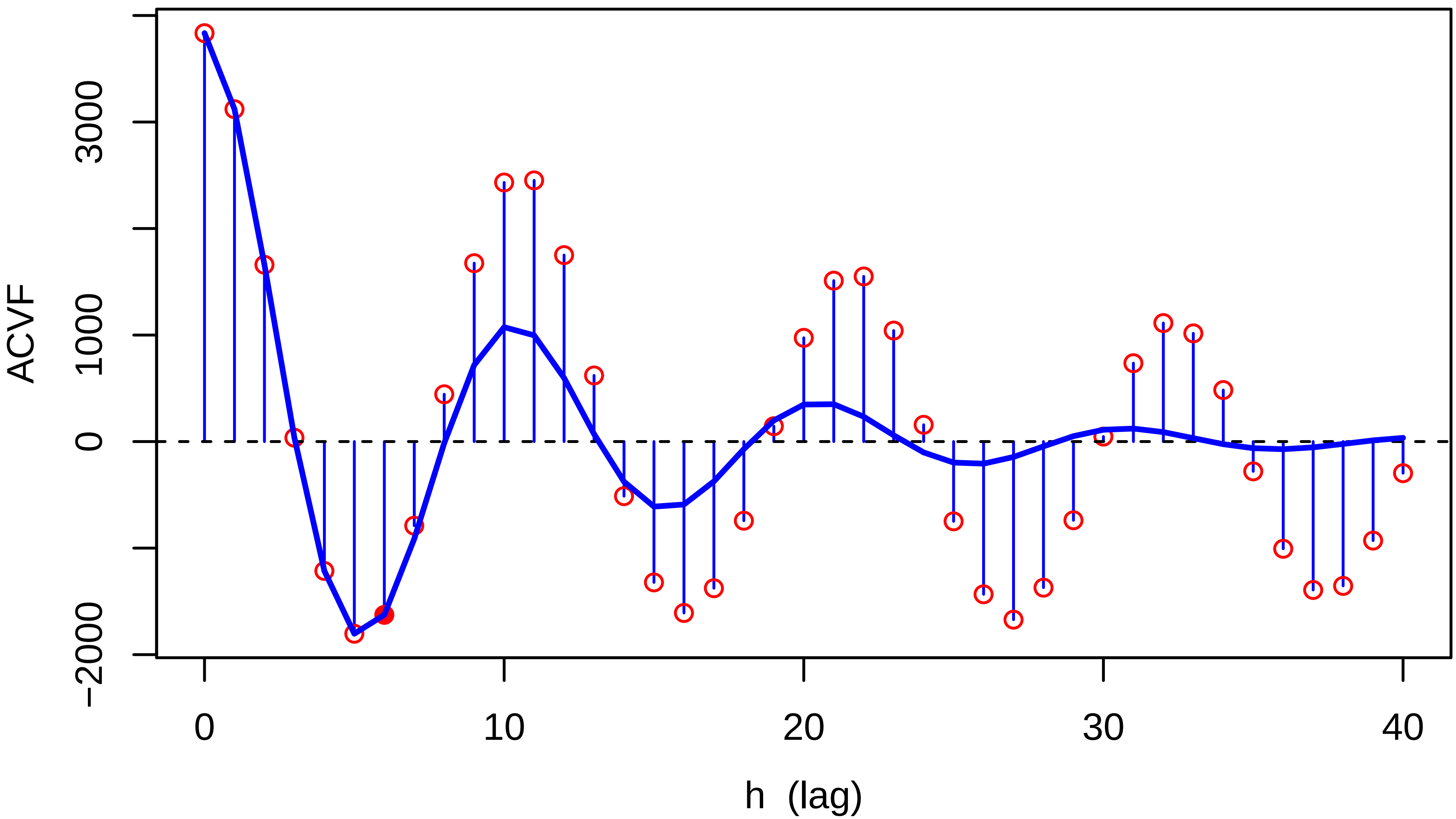
Sample and Fitted AR(2) ACVFs for Sunspots



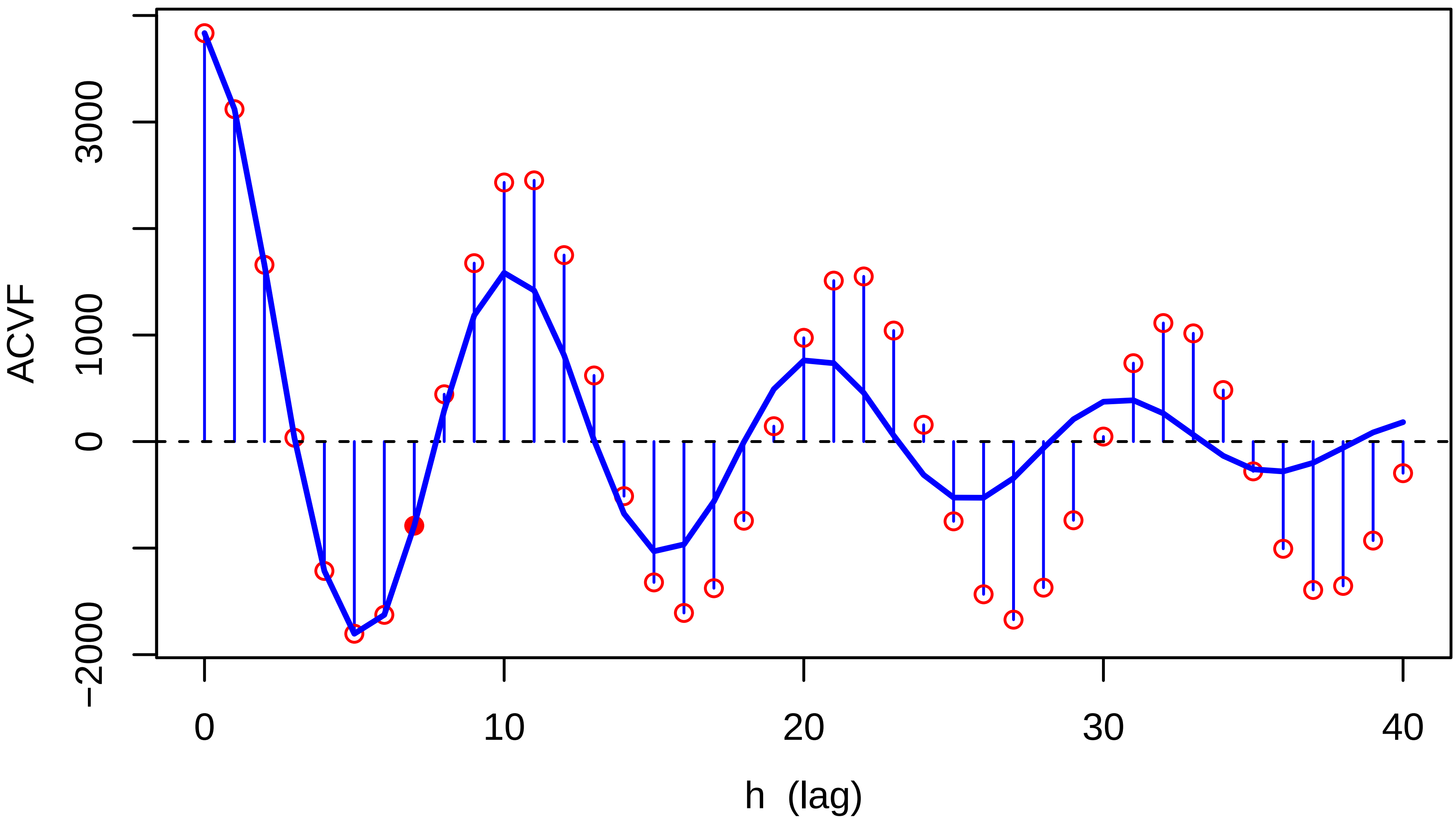


Sample and Fitted AR(3) ACVFs for Sunspots
ACVF
−2000
0
1000
3000
0
10
20
30
40
h (lag)

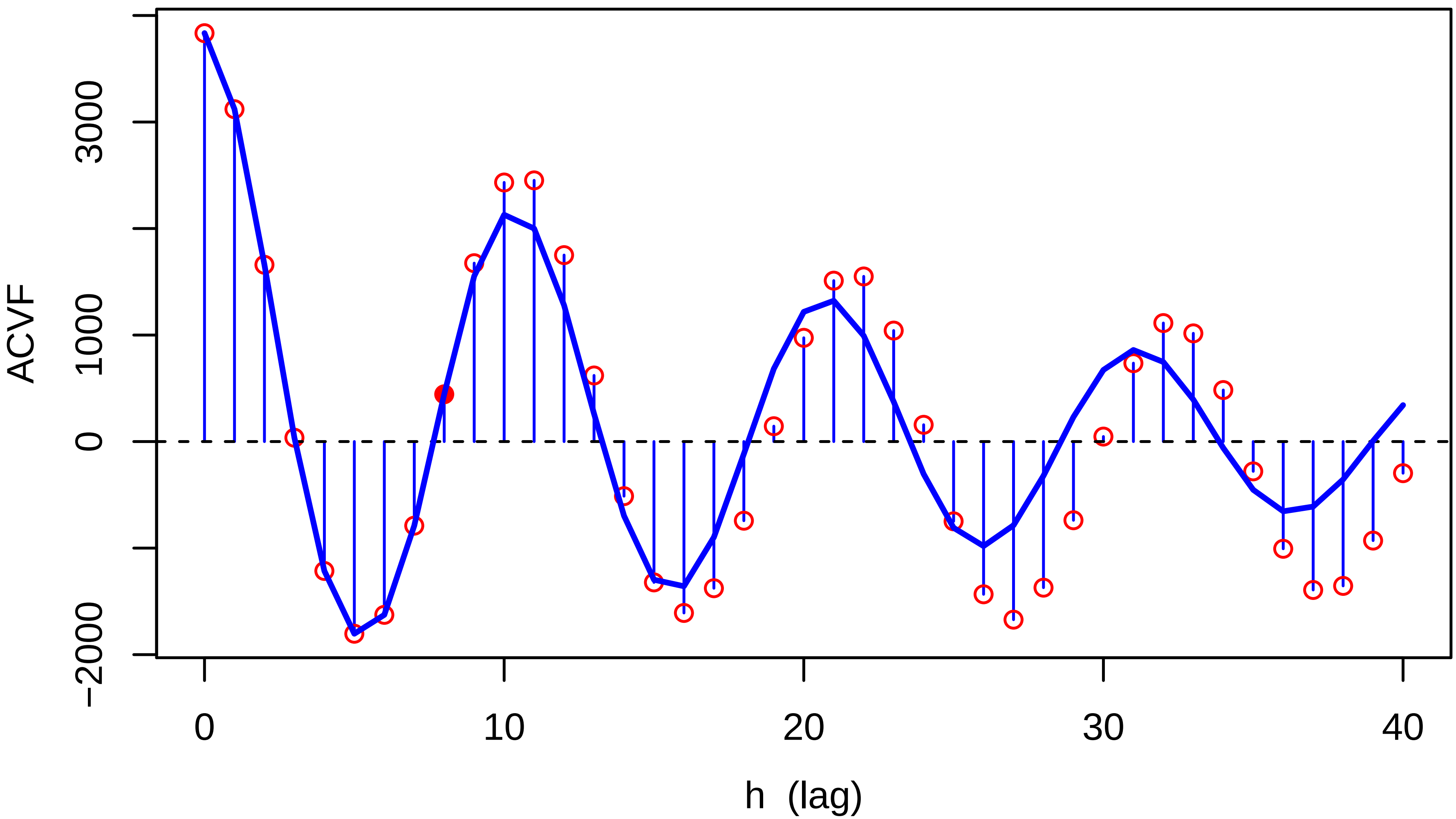
Sample and Fitted AR(6) ACVFs for Sunspots



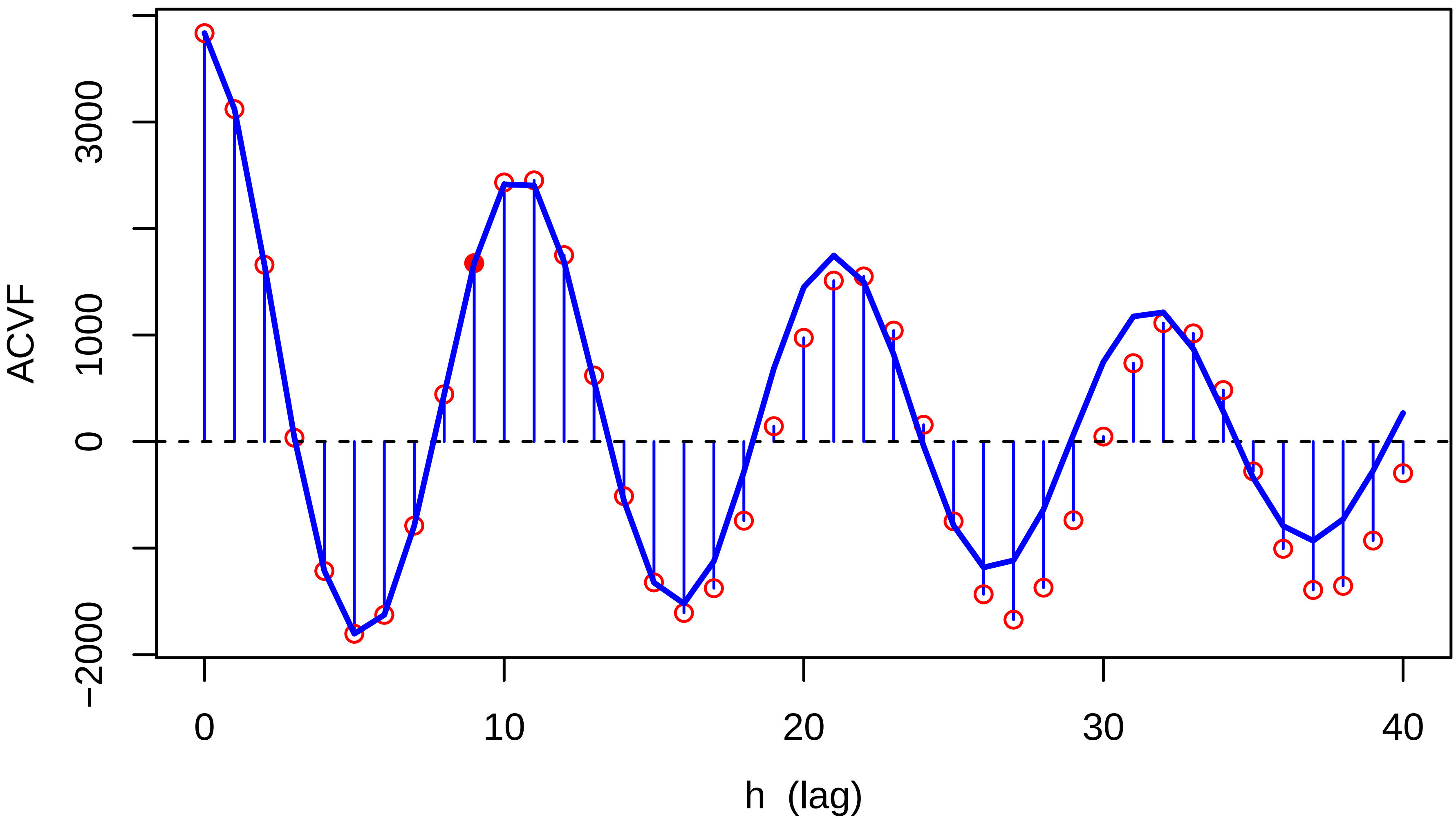
Sample and Fitted AR(7) ACVFs for Sunspots



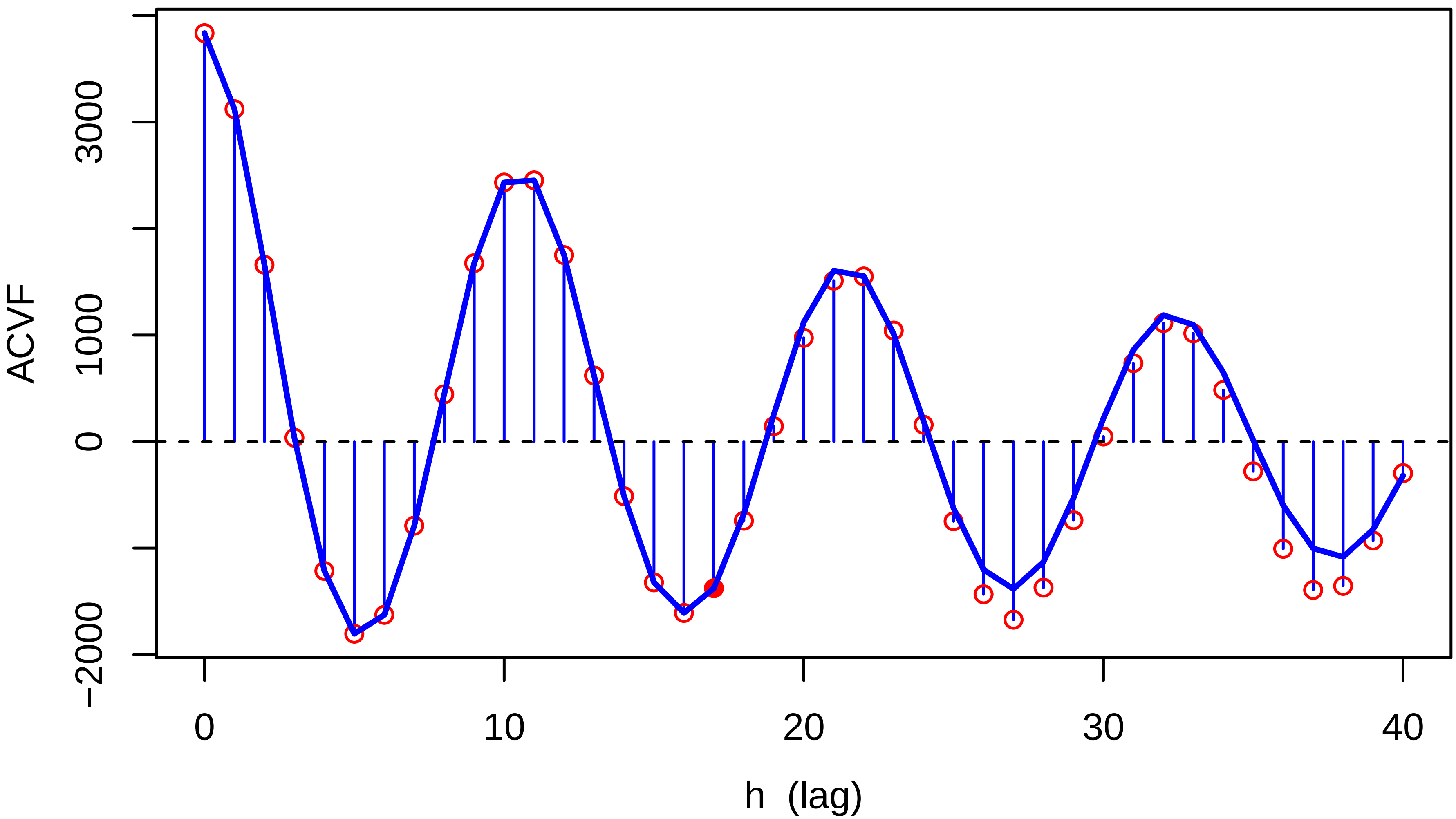
Sample and Fitted AR(8) ACVFs for Sunspots



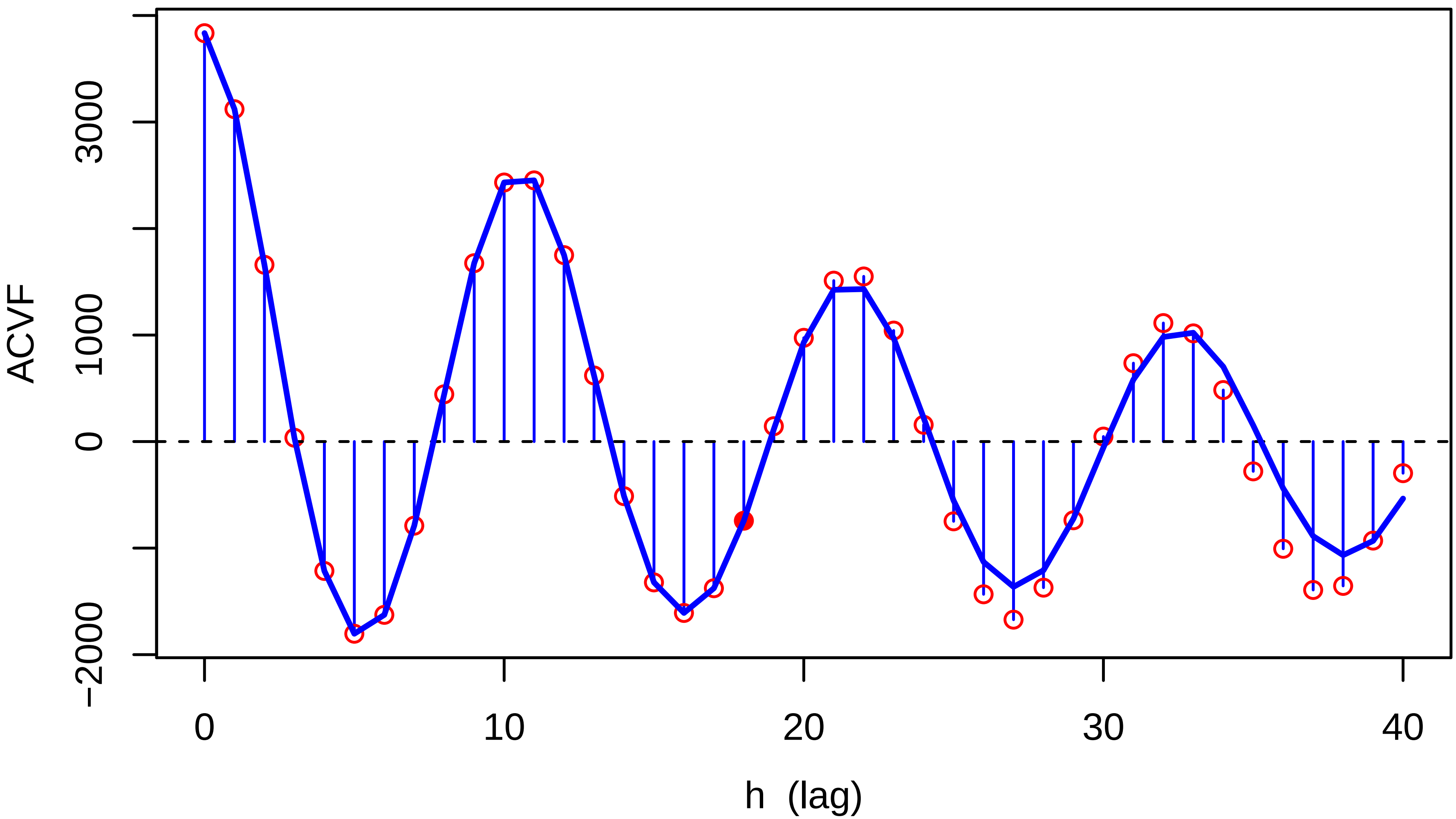
Sample and Fitted AR(9) ACVFs for Sunspots



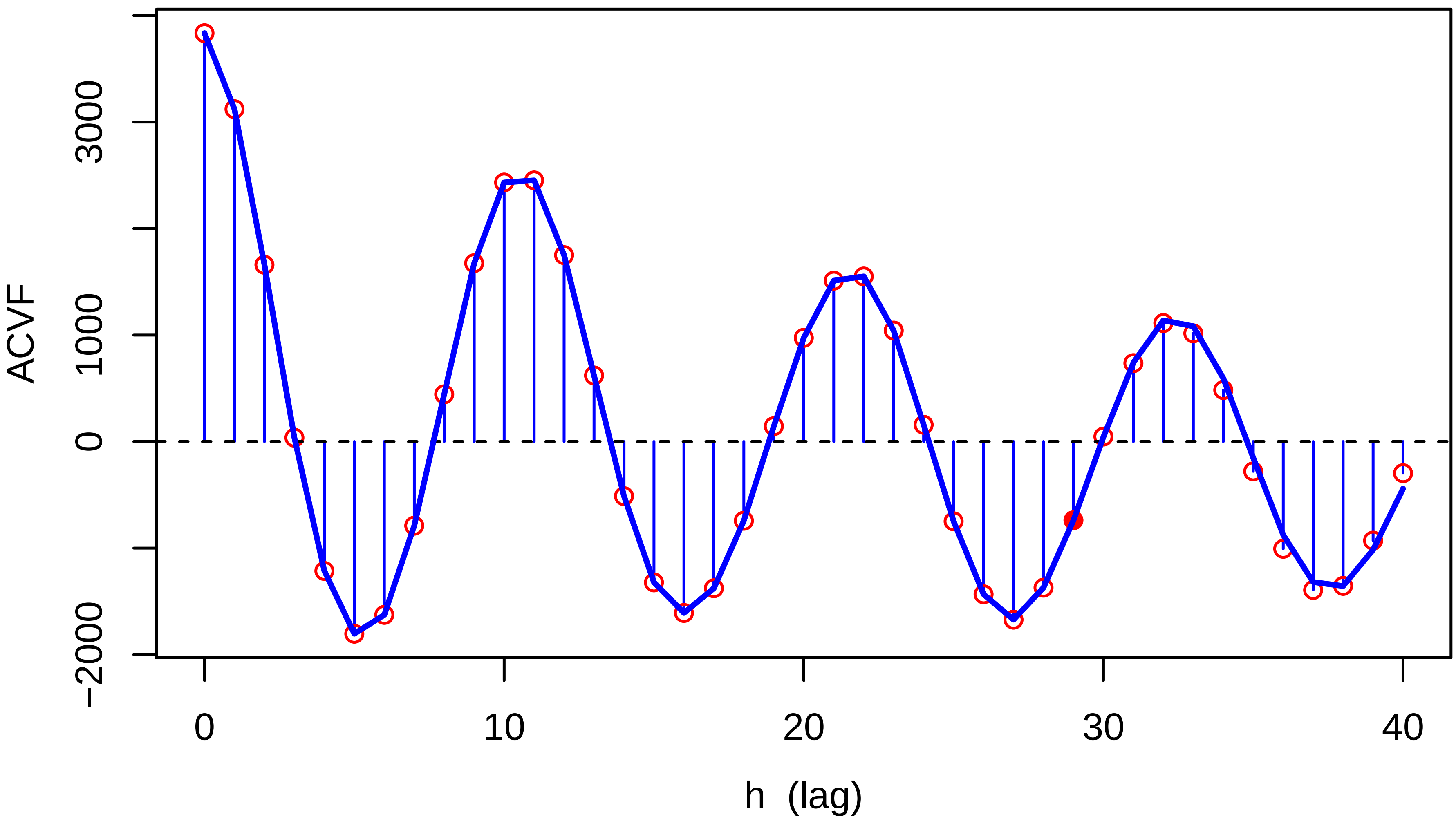
Sample and Fitted AR(17) ACVFs for Sunspots



Sample and Fitted AR(18) ACVFs for Sunspots



Sample and Fitted AR(29) ACVFs for Sunspots



Yule–Walker Estimation: VII

- distribution of Y–W estimators $\hat{\boldsymbol{\phi}}$ is approximately multivariate normal with mean $\boldsymbol{\phi}$ & covariance $\sigma^2\Gamma_p^{-1}/n$ for large n
- large sample distribution of ML estimators is the same
- don't even need to worry about inverting Γ_p: can show that
$$\sigma^2\Gamma_p^{-1} = A'A - B'B = AA' - BB',$$
where A and B are $p \times p$ lower triangular matrices whose first columns are, respectively,
$$\begin{bmatrix} 1 \\ -\phi_1 \\ \vdots \\ -\phi_{p-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \phi_p \\ \phi_{p-1} \\ \vdots \\ \phi_1 \end{bmatrix};$$
A has the same element along any given diagonal; and B has a similar structure (sometimes referred to as a *Toeplitz* structure)

Confidence Intervals and Regions for ϕ

- can use large sample distribution to get approximate confidence intervals for individual ϕ_j's or confidence region for vector $\boldsymbol{\phi}$
- approximate 95% confidence interval for ϕ_j given by

$$\left[\hat{\phi}_j - 1.96\,\frac{\hat{v}_{j,j}^{1/2}}{\sqrt{n}}, \hat{\phi}_j + 1.96\,\frac{\hat{v}_{j,j}^{1/2}}{\sqrt{n}}\right],$$

  where $\hat{v}_{j,j}$ is jth diagonal element of $\hat{\sigma}^2\hat{\Gamma}_p^{-1}$

- letting $\chi^2_{0.95}(p)$ denote 95% quantile of chi-squared distribution with p degrees of freedom, approximate 95% confidence region for $\boldsymbol{\phi}$ is the set of all $\boldsymbol{\phi}$'s such that

$$(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})'\hat{\Gamma}_p(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \leq \chi^2_{0.95}(p)\frac{\hat{\sigma}^2}{n}$$

Yule–Walker Estimation and Order Selection: I

- when Y–W is used to estimate coefficients for AR(h) model
$$X_t - \phi_1 X_{t-1} - \cdots - \phi_h X_{t-h} = Z_t,$$
estimate $\hat{\phi}_h$ is same as $\hat{\phi}_{h,h}$ (hth member of sample PACF)
- as noted before, large sample theory suggests that $\hat{\phi}_{h,h}$ is approximately $\mathcal{N}(0, 1/n)$ for $h > p$ (the true AR model order)
- given estimates of $\hat{\phi}_{h,h}$ out to some maximum order, say H, Brockwell & Davis suggest setting p to be smallest m such that $|\hat{\phi}_{h,h}| < 1.96/\sqrt{n}$ for $m < h \leq H$
- obvious danger: sampling variability might result in p being set too high
  - with $H = 40$ in sunspot example, would select $p = 29$, which might not be a reasonable choice

Yule–Walker Estimation and Order Selection: II

- another approach is to select order that minimizes AICC statistic (biased-corrected version of Akaike's information criterion):

$$\text{AICC} = -2\ln\left(L(\boldsymbol{\phi}, S(\boldsymbol{\phi})/n)\right) + \frac{2(p+1)n}{n-p-2},$$

  where L is Gaussian likelihood function, and $S(\boldsymbol{\phi})$ is defined below

- given Gaussian AR(p) time series $\boldsymbol{X}_n = [X_n, \ldots, X_1]'$ with mean zero and covariance matrix Γ_n (implicitly dependent on $\boldsymbol{\phi}$ & σ^2), can write

$$L(\Gamma_n) = (2\pi)^{-n/2}(\det \Gamma_n)^{-1/2} \exp\left(-\tfrac{1}{2}\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n\right)$$

  and hence

$$-2\ln\left(L(\Gamma_n)\right) = n\ln(2\pi) + \ln(\det \Gamma_n) + \boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n$$

Yule–Walker Estimation and Order Selection: III

- when considering ML estimation later on, will argue that
$$-2\ln\left(L(\Gamma_n)\right) = n\ln(2\pi) + \ln\left(\det\Gamma_n\right) + \boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n$$
can be rewritten in AR(p) case as
$$-2\ln L(\boldsymbol{\phi},\sigma^2) = n\ln(2\pi\sigma^2) + \sum_{j=0}^{p-1}\ln(r_j) + \frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{(X_j-\widehat{X}_j)^2}{r_{j-1}},$$
where $r_j \stackrel{\text{def}}{=} v_j/\sigma^2$ & hence $v_j = r_j\sigma^2$ (note: $r_j = 1$ for $j \geq p$)
- dependence on $\boldsymbol{\phi}$ is through v_j's and coefficients determining $\widehat{X}_j$ (can get these from $\boldsymbol{\phi}$ using step-down L–D recursions)
- can remove σ^2 by replacing it with $S(\boldsymbol{\phi})/n$, where
$$S(\boldsymbol{\phi}) = \sum_{j=1}^{n}\frac{(X_j-\widehat{X}_j)^2}{r_{j-1}}$$

Yule–Walker Estimation and Order Selection: IV

- with removal of σ^2, AICC statistic becomes

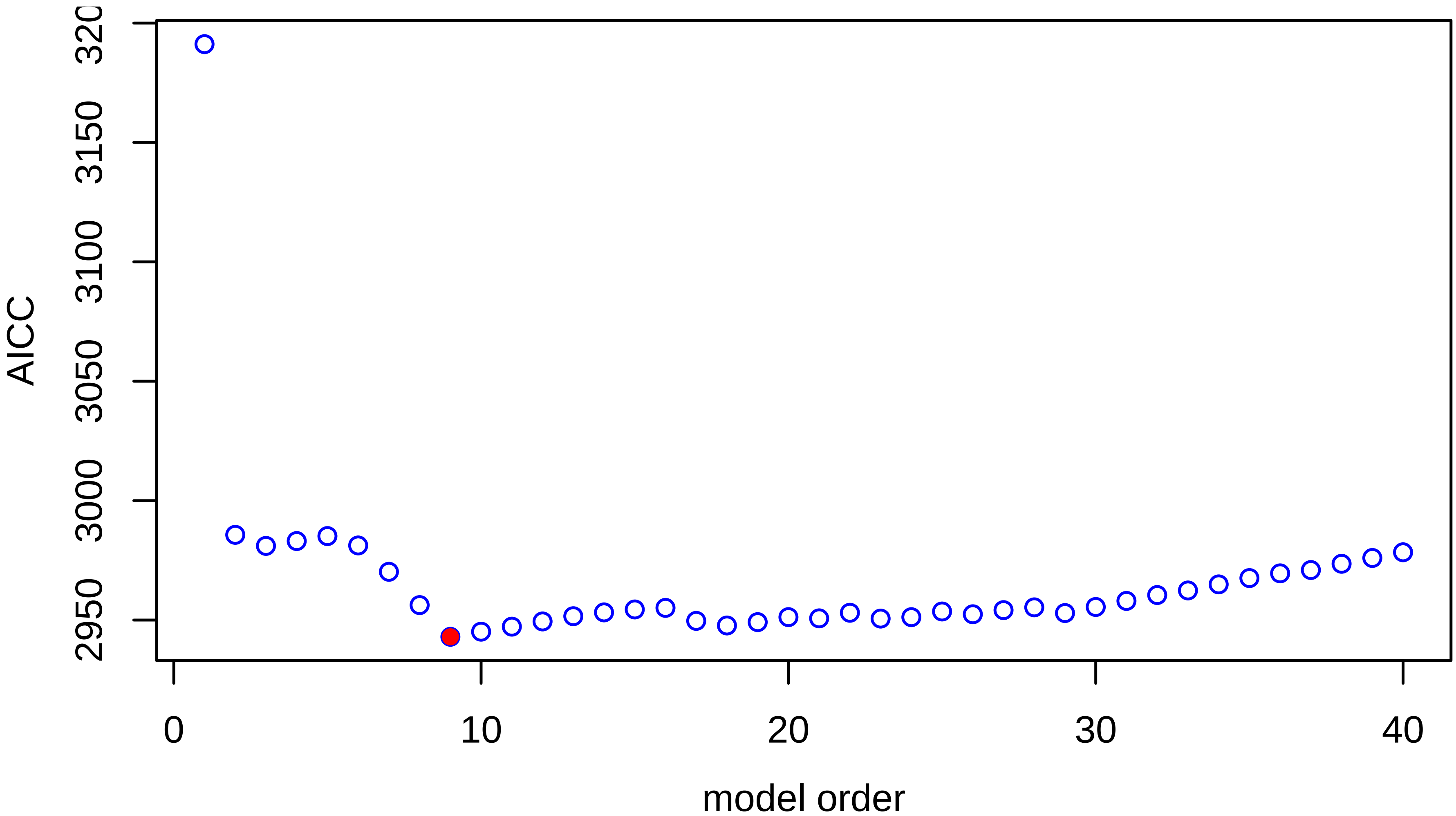
$$\text{AICC} = C_n + n \ln \left(\sum_{j=1}^{n} \frac{(X_j - \widehat{X}_j)^2}{r_{j-1}} \right) + \sum_{j=0}^{p-1} \ln(r_j) + \frac{2(p+1)n}{n-p-2},$$

  where

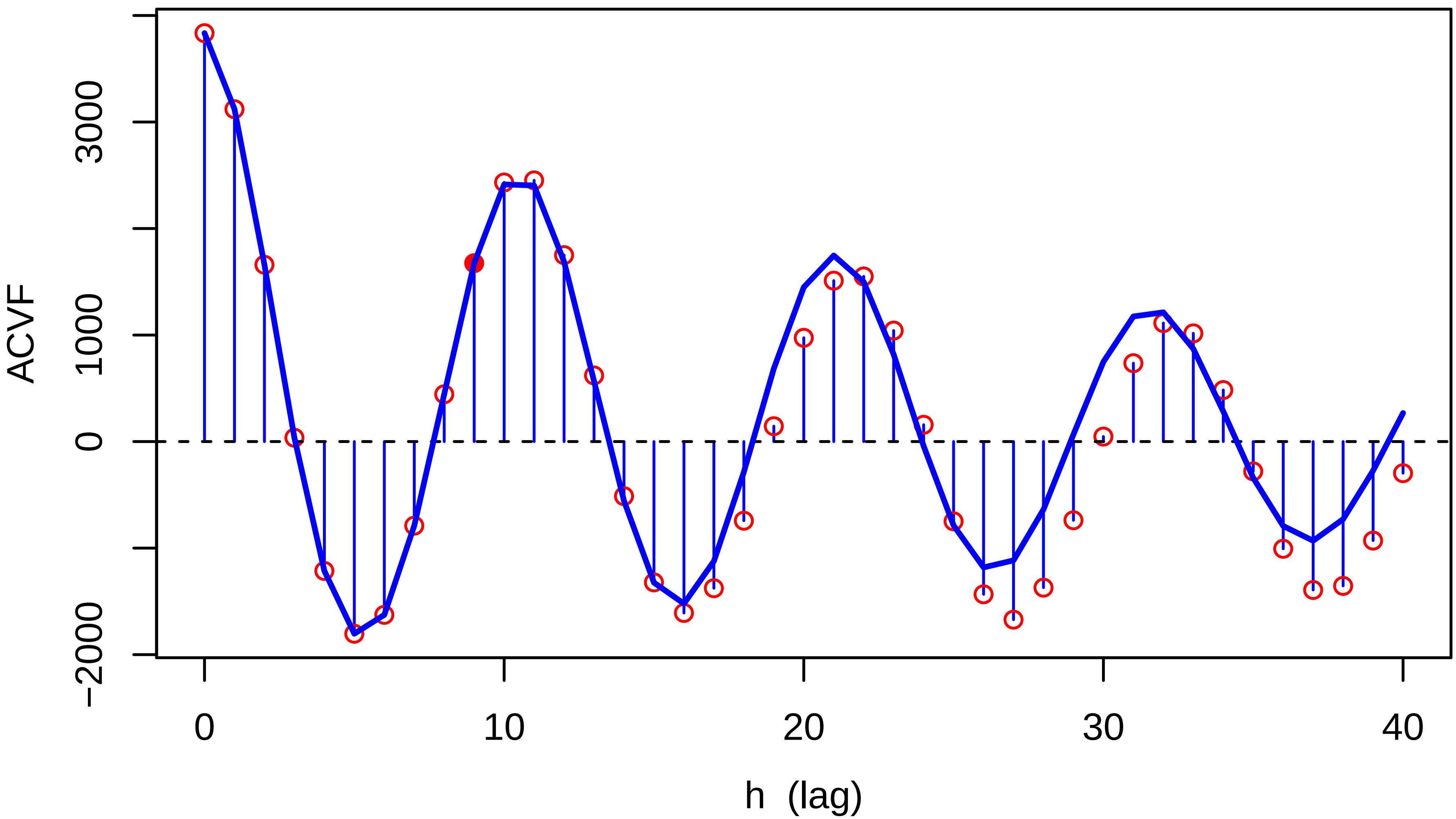
$$C_n \stackrel{\text{def}}{=} n + n \ln(2\pi/n)$$

- note: will discuss other order selection statistics (BIC etc.) later
- let's see what order the AICC picks out for sunspot series

AICC for Sunspots



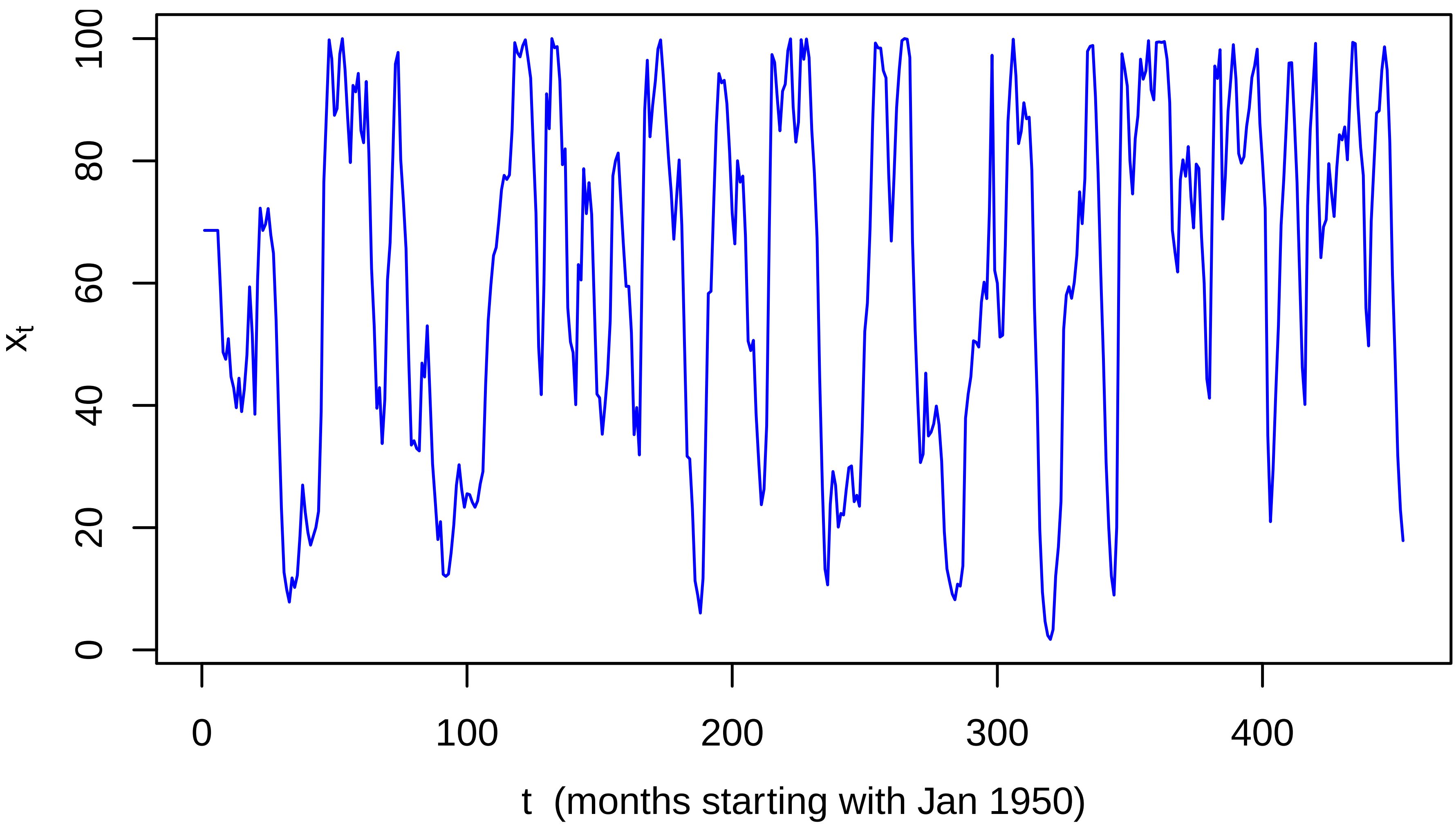
Sample and Fitted AR(9) ACVFs for Sunspots

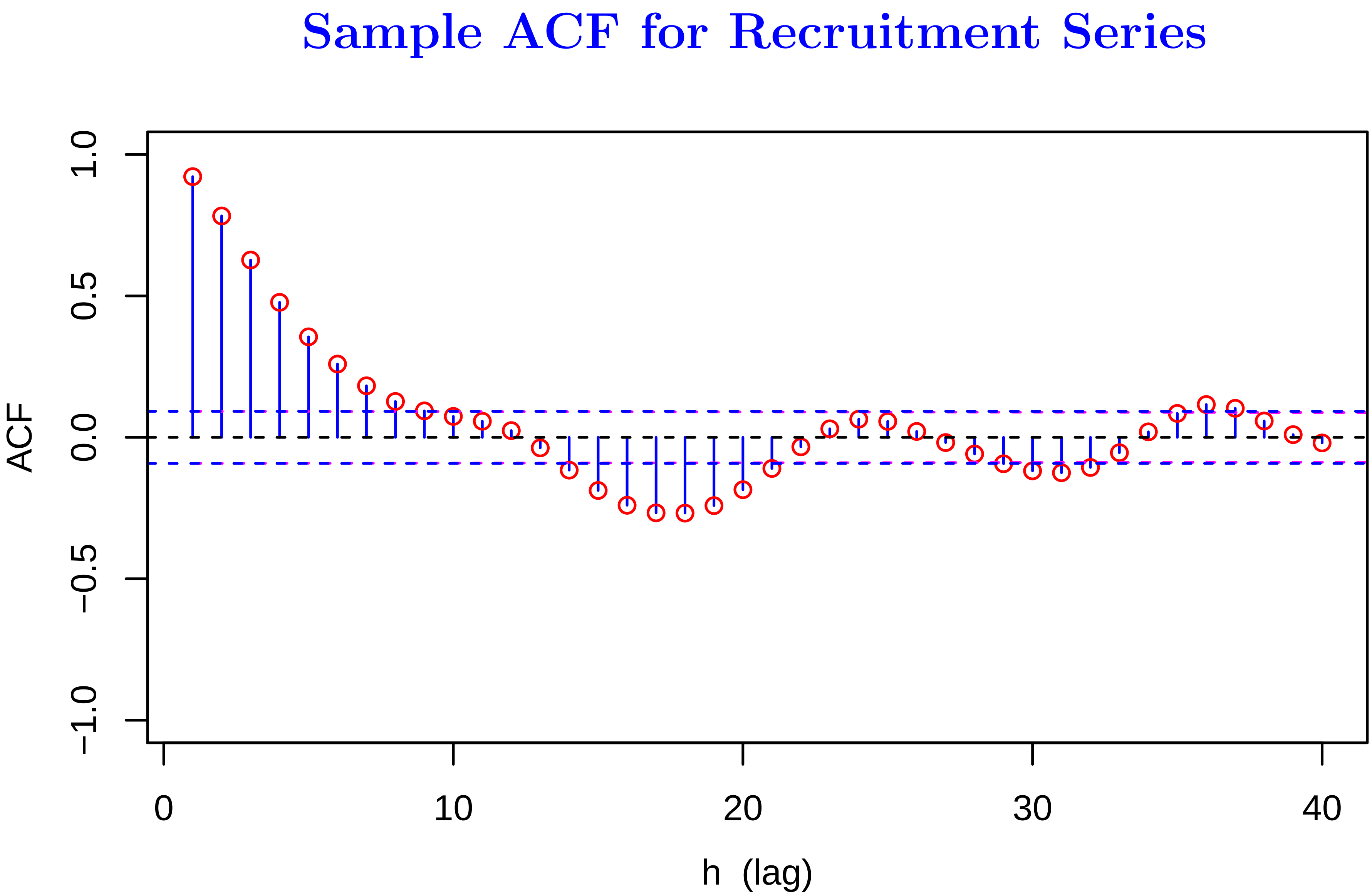


Example – Recruitment Time Series: I

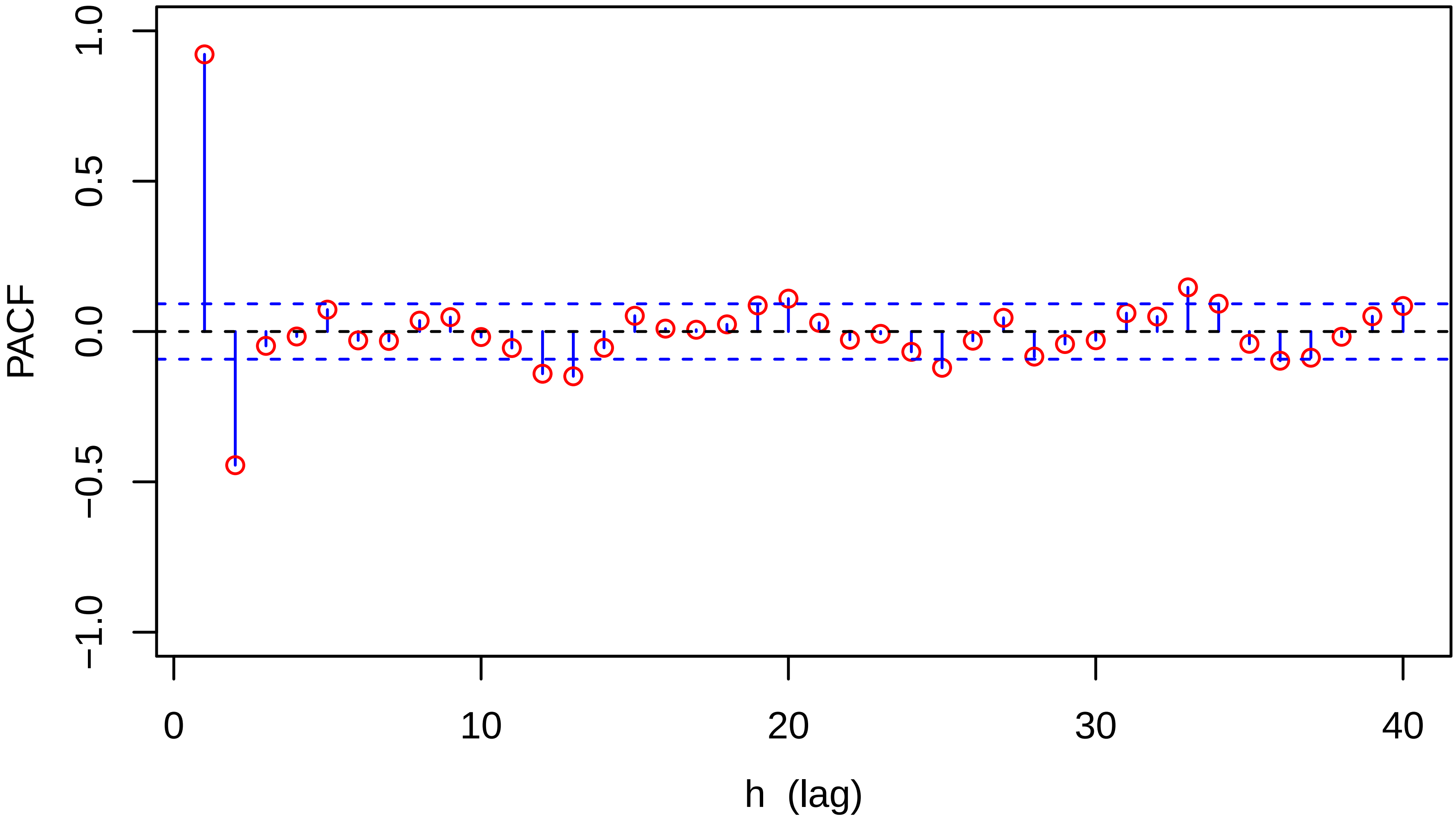
- monthly measure of number of new fish entering Pacific Ocean (453 months covering 1950–87; Shumway & Stoffer got it from Roy Mendelssohn, NOAA/PFEL, who got it from Pierre Kleiber NOAA/NMFS, who generated measures using a model . . .)

Recruitment Time Series (1950–1987)

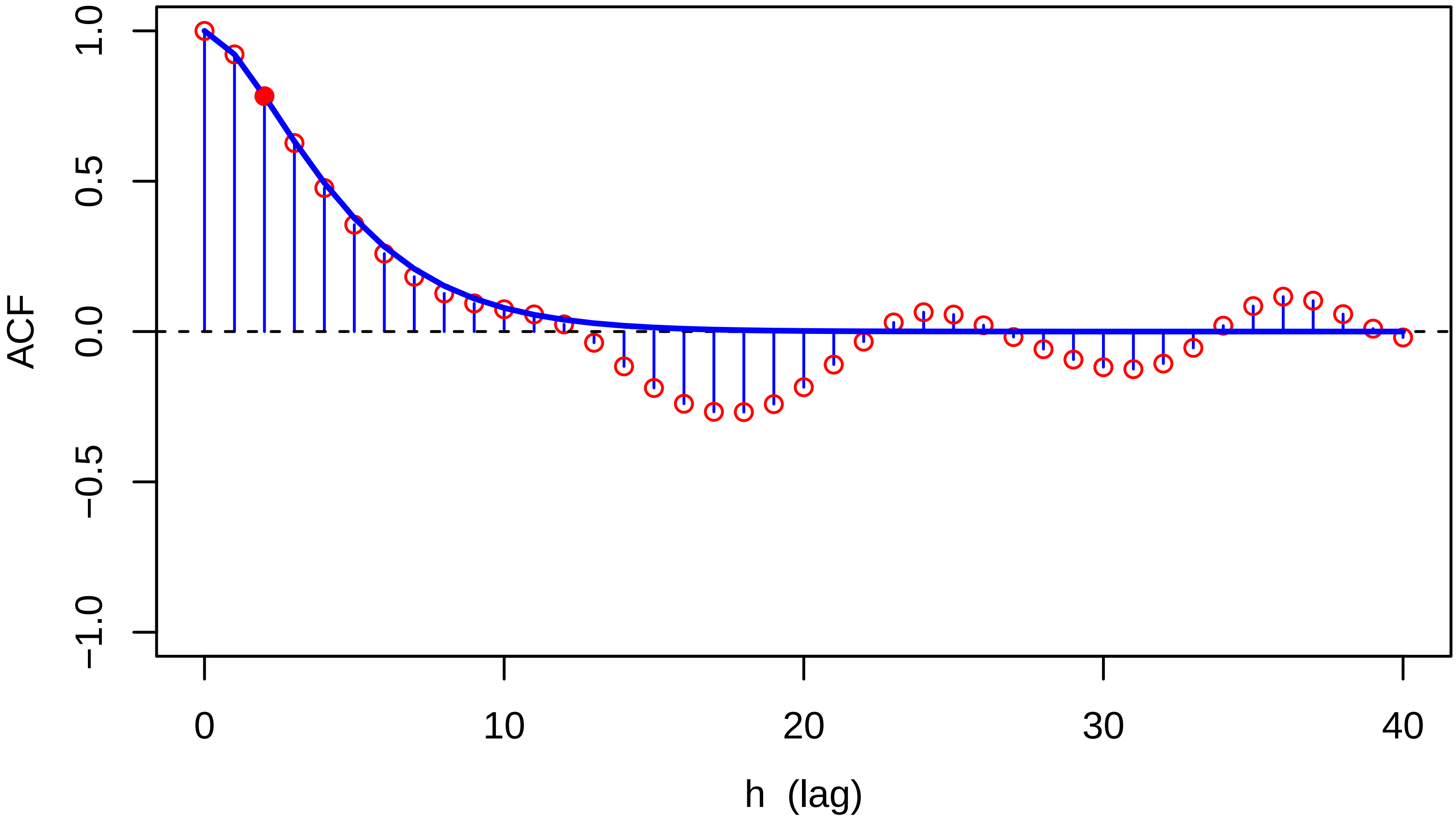



Sample ACF for Recruitment Series
ACF
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
h (lag)

Sample PACF for Recruitment Series



Sample & Fitted AR(2) ACFs for Recruitment Series



Example – Recruitment Time Series: II

- for AR(2) model, Y–W estimates are

$$\hat{\boldsymbol{\phi}} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} \doteq \begin{bmatrix} 1.3316 \\ -0.4445 \end{bmatrix} \text{ and } \hat{\sigma}^2 \doteq 94.171$$

(note: `R` function **`ar`** gives $\hat{\sigma}^2 \doteq 94.799$ … hmmm)

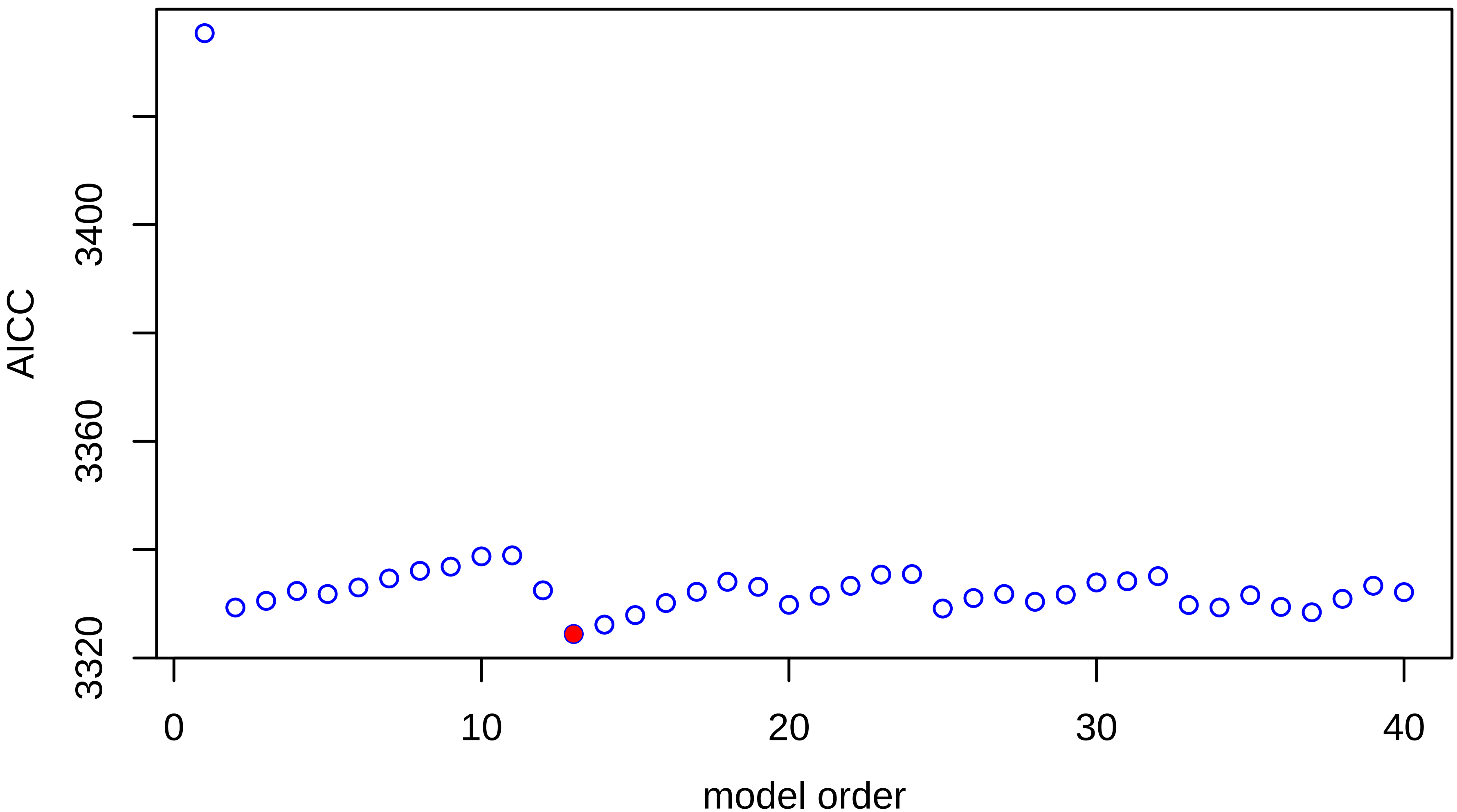
- using large sample approximation that $\hat{\boldsymbol{\phi}}$ is multivariate normal with mean $\boldsymbol{\phi}$ and covariance $\sigma^2\Gamma_2^{-1}$, can get 95% confidence intervals (CIs) and regions based upon

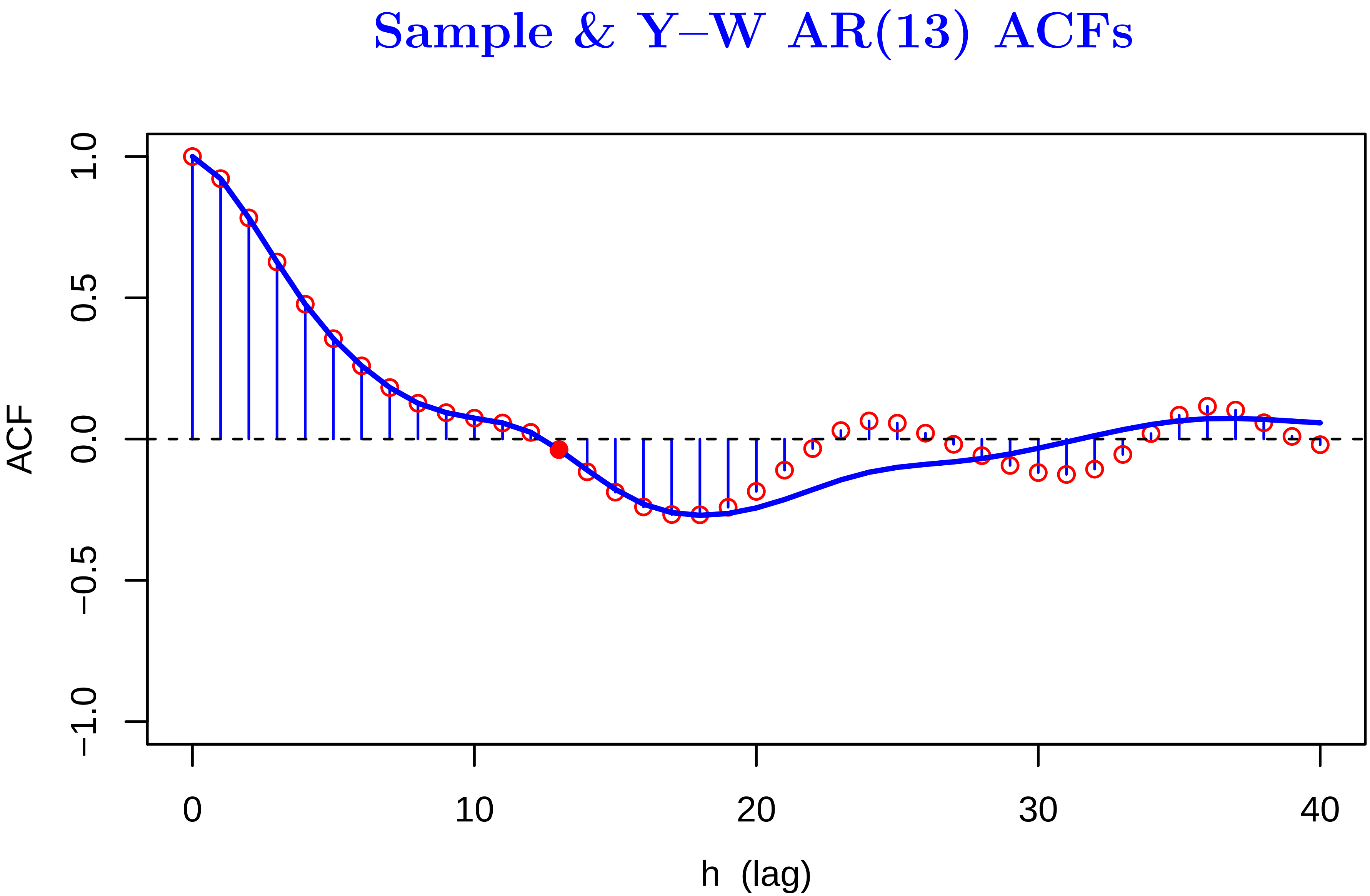
$$\hat{\sigma}^2\hat{\Gamma}_2^{-1} = \hat{\sigma}^2 \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) \end{bmatrix}^{-1} = \begin{bmatrix} \hat{v}_{1,1} & \hat{v}_{1,2} \\ \hat{v}_{2,1} & \hat{v}_{2,2} \end{bmatrix} \doteq \begin{bmatrix} 0.8024 & -0.7396 \\ -0.7396 & 0.8024 \end{bmatrix}$$

- using $\left[\hat{\phi}_j - 1.96\,\hat{v}_{j,j}^{1/2}/\sqrt{n}, \hat{\phi}_j + 1.96\,\hat{v}_{j,j}^{1/2}/\sqrt{n}\right]$ yields 95% CIs

$$[1.2491, 1.4141] \text{ for } \phi_1 \text{ and } [-0.5270, -0.3621] \text{ for } \phi_2$$

AICC for Recruitment Series (Y–W)




Sample & Y–W AR(13) ACFs
ACF
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
h (lag)

Burg's Algorithm: I

- Y–W estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ is based on L–D recursions with $\gamma(h)$ replaced by $\hat{\gamma}(h)$
- given $\hat{\boldsymbol{\phi}}_{k-1}$ & $\hat{v}_{k-1}$, recursion gives us $\hat{\boldsymbol{\phi}}_k$ & $\hat{v}_k$ via 3 steps
  1. get kth order partial autocorrelation:
  $$\hat{\phi}_{k,k} = \frac{\hat{\gamma}(k) - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\hat{\gamma}(k-j)}{v_{k-1}}$$
  2. get remaining $\hat{\phi}_{k,j}$'s:
  $$\begin{bmatrix} \hat{\phi}_{k,1} \\ \vdots \\ \hat{\phi}_{k,k-1} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{k-1,1} \\ \vdots \\ \hat{\phi}_{k-1,k-1} \end{bmatrix} - \hat{\phi}_{k,k} \begin{bmatrix} \hat{\phi}_{k-1,k-1} \\ \vdots \\ \hat{\phi}_{k-1,1} \end{bmatrix}$$
  3. get kth order MSE: $\hat{v}_k = \hat{v}_{k-1}(1 - \hat{\phi}_{k,k}^2)$

Burg's Algorithm: II

- when $k = p$, get Y–W estimators $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_p$ and $\hat{\sigma}^2 = \hat{v}_p$
- start procedure by setting

$$\hat{\boldsymbol{\phi}}_1 = [\hat{\phi}_{1,1}] = \left[\frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}\right] \text{ and } \hat{v}_1 = \hat{\gamma}(0)(1 - \hat{\phi}_{1,1}^2)$$

- sample ACVF comes into play in forming $\hat{\boldsymbol{\phi}}_1$ and in 1st step of L–D recursions, but not in 2nd and 3rd steps
- sample ACVF just used to get PACF estimates $\hat{\phi}_{1,1}, \ldots, \hat{\phi}_{p,p}$
- kth component of PACF is a correlation coefficient:

$$\phi_{k,k} = \text{corr}\left\{X_k - \widehat{X}_k, X_0 - \widehat{X}_{0|k-1}\right\}$$

($\widehat{X}_{0|k-1}$ is best linear predictor of X_0 given $X_1, \ldots, X_{k-1}$)

- Burg's algorithm is based on estimating $\phi_{k,k}$ in keeping with the above rather than via sample ACVF

Burg's Algorithm: III

- let $\bar{\boldsymbol{\phi}}_{k-1} = [\bar{\phi}_{k-1,1}, \ldots, \bar{\phi}_{k-1,k-1}]'$ be Burg estimator of coefficients for AR$(k-1)$ process based on $X_1, \ldots, X_n$
- calculate forward & backward observed innovations:

$$\overrightarrow{U}_t(k-1) \stackrel{\text{def}}{=} X_t - \sum_{j=1}^{k-1} \bar{\phi}_{k-1,j} X_{t-j}, \quad k \le t \le n$$

$$\overleftarrow{U}_{t-k}(k-1) \stackrel{\text{def}}{=} X_{t-k} - \sum_{j=1}^{k-1} \bar{\phi}_{k-1,j} X_{t-k+j}, \quad k+1 \le t \le n+1$$

- can show that, for *any* estimator $\bar{\phi}_{k,k}$ with $\bar{\phi}_{k,1}, \ldots, \bar{\phi}_{k,k-1}$ generated by step 2 of L–D, have, for $k+1 \le t \le n$

$$\overrightarrow{U}_t(k) = \overrightarrow{U}_t(k-1) - \bar{\phi}_{k,k} \overleftarrow{U}_{t-k}(k-1)$$
$$\overleftarrow{U}_{t-k}(k) = \overleftarrow{U}_{t-k}(k-1) - \bar{\phi}_{k,k} \overrightarrow{U}_t(k-1)$$

Burg's Algorithm: IV

- Burg's idea: choose $\bar{\phi}_{k,k}$ that minimizes

$$\mathrm{SS}_k(\bar{\phi}_{k,k}) \stackrel{\mathrm{def}}{=} \sum_{t=k+1}^{n} \overrightarrow{U}_t^2(k) + \overleftarrow{U}_{t-k}^2(k)$$

- yields Burg's estimator

$$\bar{\phi}_{k,k} \stackrel{\mathrm{def}}{=} \frac{\sum_{t=k+1}^{n} \overrightarrow{U}_t(k-1)\overleftarrow{U}_{t-k}(k-1)}{\frac{1}{2}\sum_{t=k+1}^{n} \overrightarrow{U}_t^2(k-1) + \overleftarrow{U}_{t-k}^2(k-1)}$$

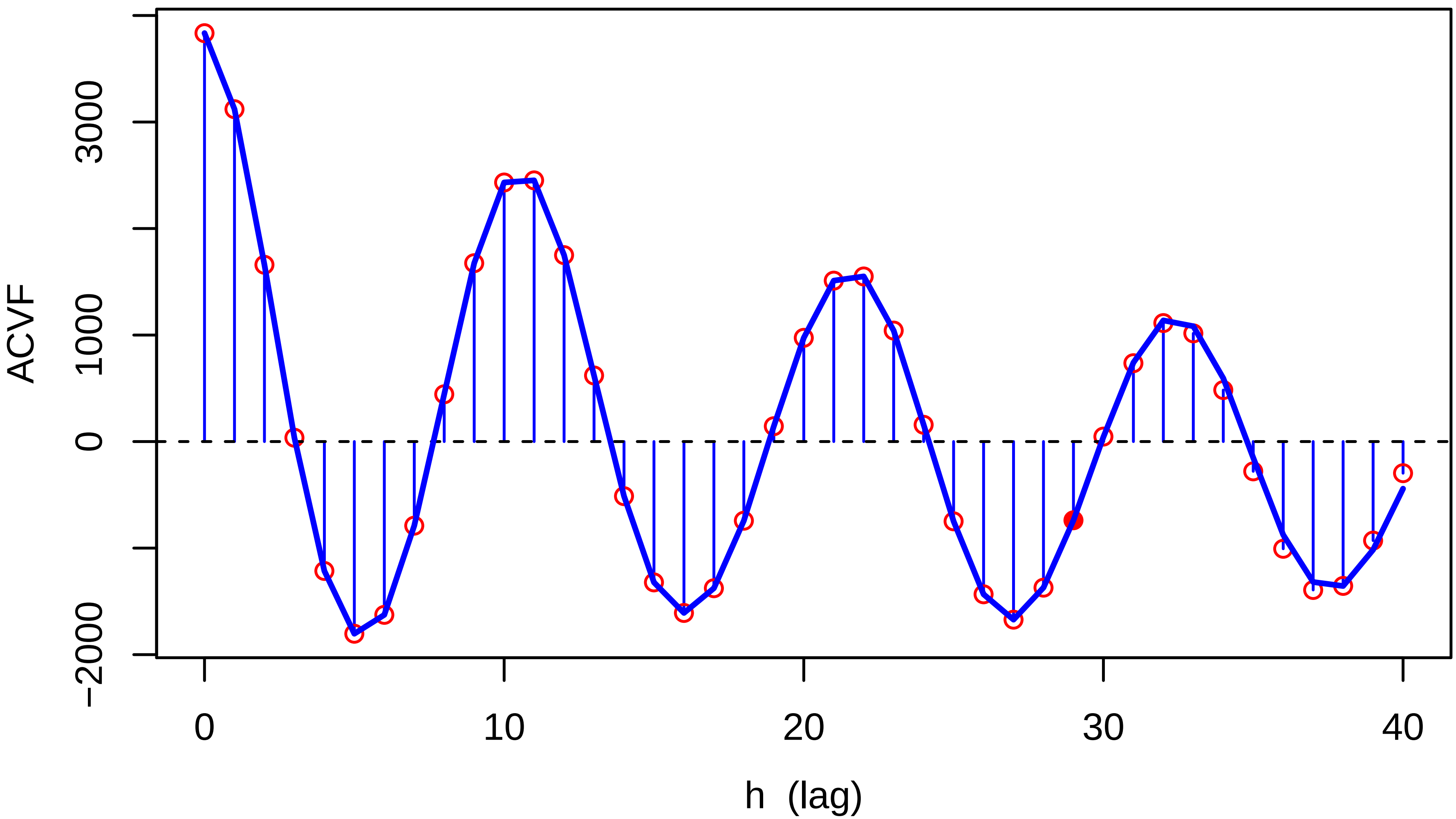
- compare above to following expression:

$$\begin{aligned}\phi_{k,k} &= \operatorname{corr}\{X_k - \widehat{X}_k, X_0 - \widehat{X}_{0|k-1}\} \\ &= \frac{\operatorname{cov}\{X_k - \widehat{X}_k, X_0 - \widehat{X}_{0|k-1}\}}{\left(\operatorname{var}\{X_k - \widehat{X}_k\}\operatorname{var}\{X_0 - \widehat{X}_{0|k-1}\}\right)^{1/2}}\end{aligned}$$

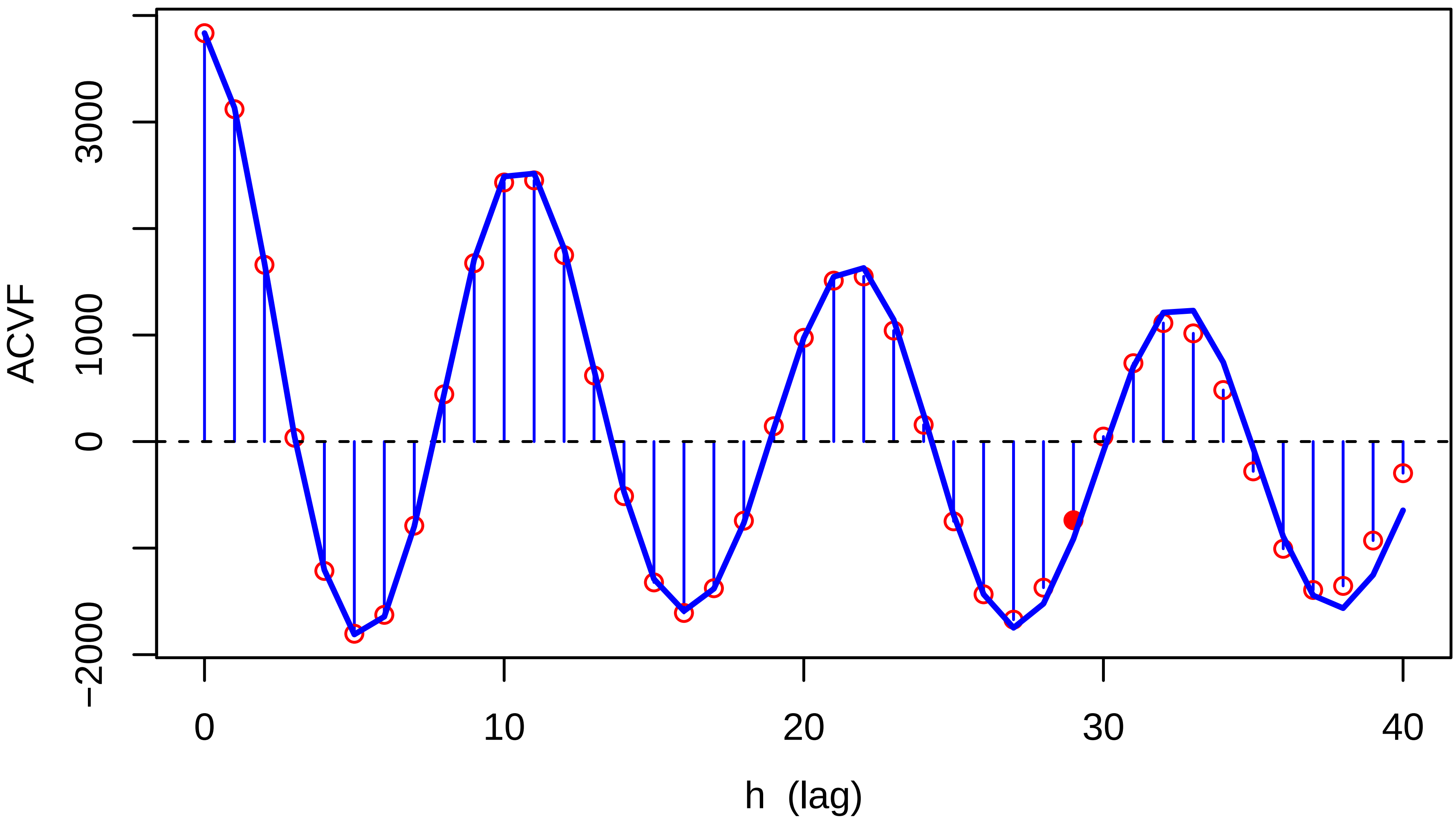
Burg's Algorithm: V

- initialize with $\overrightarrow{U}_t(0) \stackrel{\text{def}}{=} X_t$ and $\overleftarrow{U}_{t-1}(0) \stackrel{\text{def}}{=} X_{t-1}$
- guaranteed to have $|\bar{\phi}_{k,k}| \leq 1$ for all k
- if $|\bar{\phi}_{p,p}| \neq 1$, Burg estimators $\bar{\boldsymbol{\phi}} = \bar{\boldsymbol{\phi}}_p$ of coefficients $\boldsymbol{\phi}$ *always* correspond to causal (& hence stationary) AR(p) process (same is true for Y–W, but $|\hat{\phi}_p| = 1$ can't happen with Y–W)
- large sample distribution for Burg same as for Y–W and ML, but Monte Carlo studies show Burg outperforming Y–W
- unlike Y–W, AR(p) model fitted via Burg's algorithm has theoretical ACVF that need *not* be identical to sample ACVF at lags 0, 1, . . . , p, as another visit to sunspot time series shows

Sample and Y–W AR(29) ACVFs for Sunspots



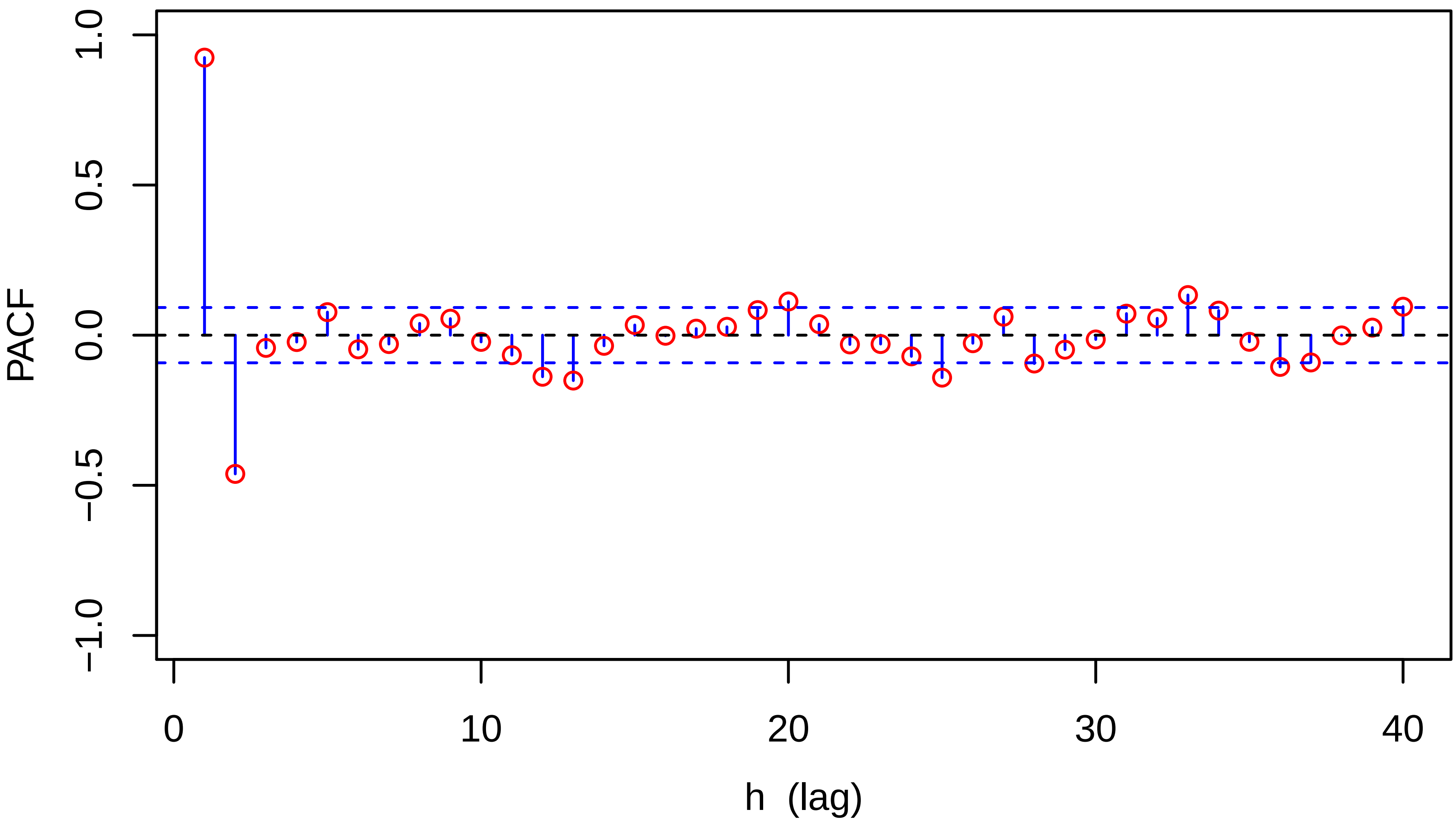
Sample and Burg AR(29) ACVFs for Sunspots



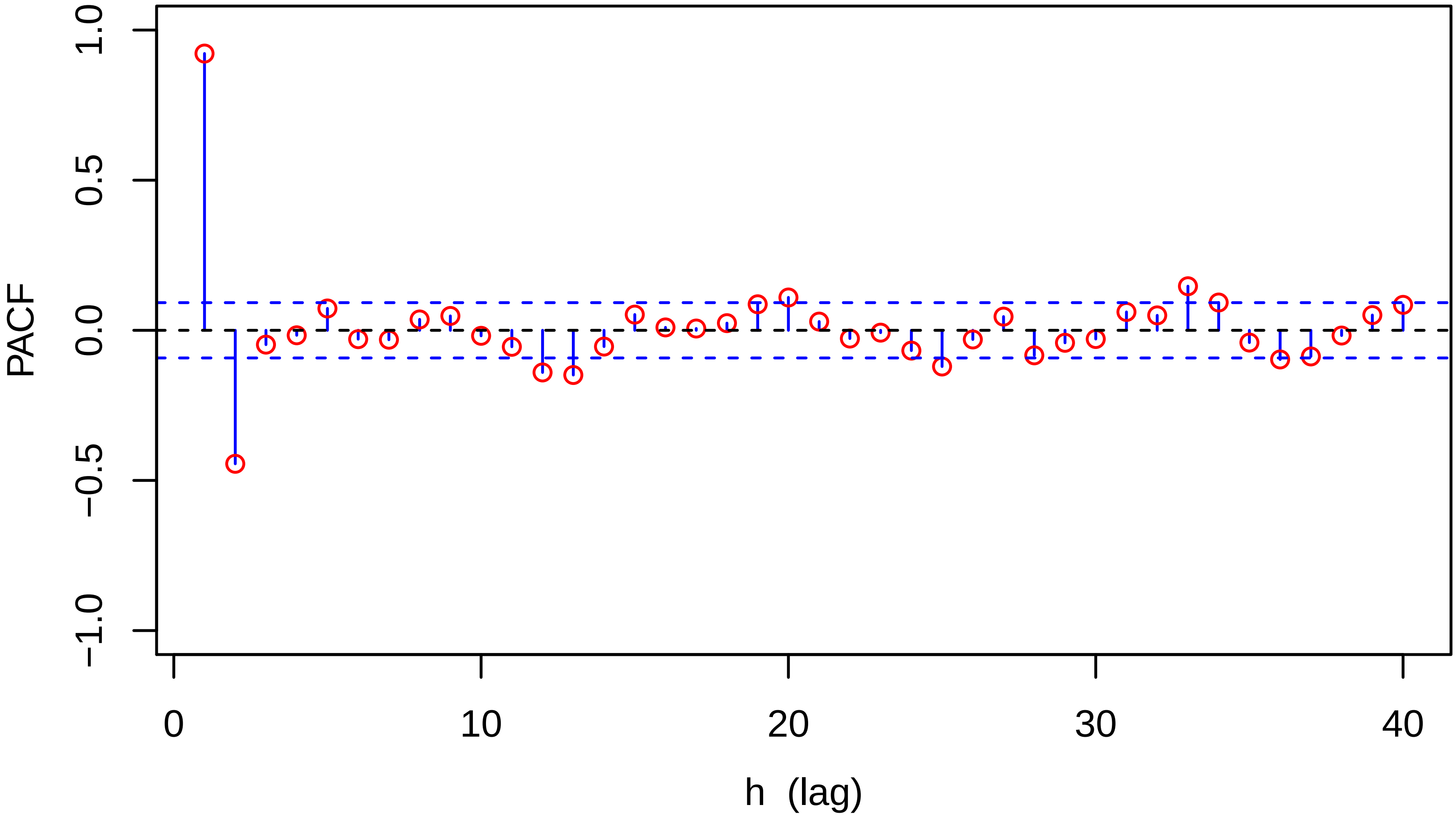
Example – Recruitment Time Series: III

- reconsider recruitment time series, this time using Burg's algorithm to get estimate of PACF that is an alternative to sample PACF (latter is based on Y–W)

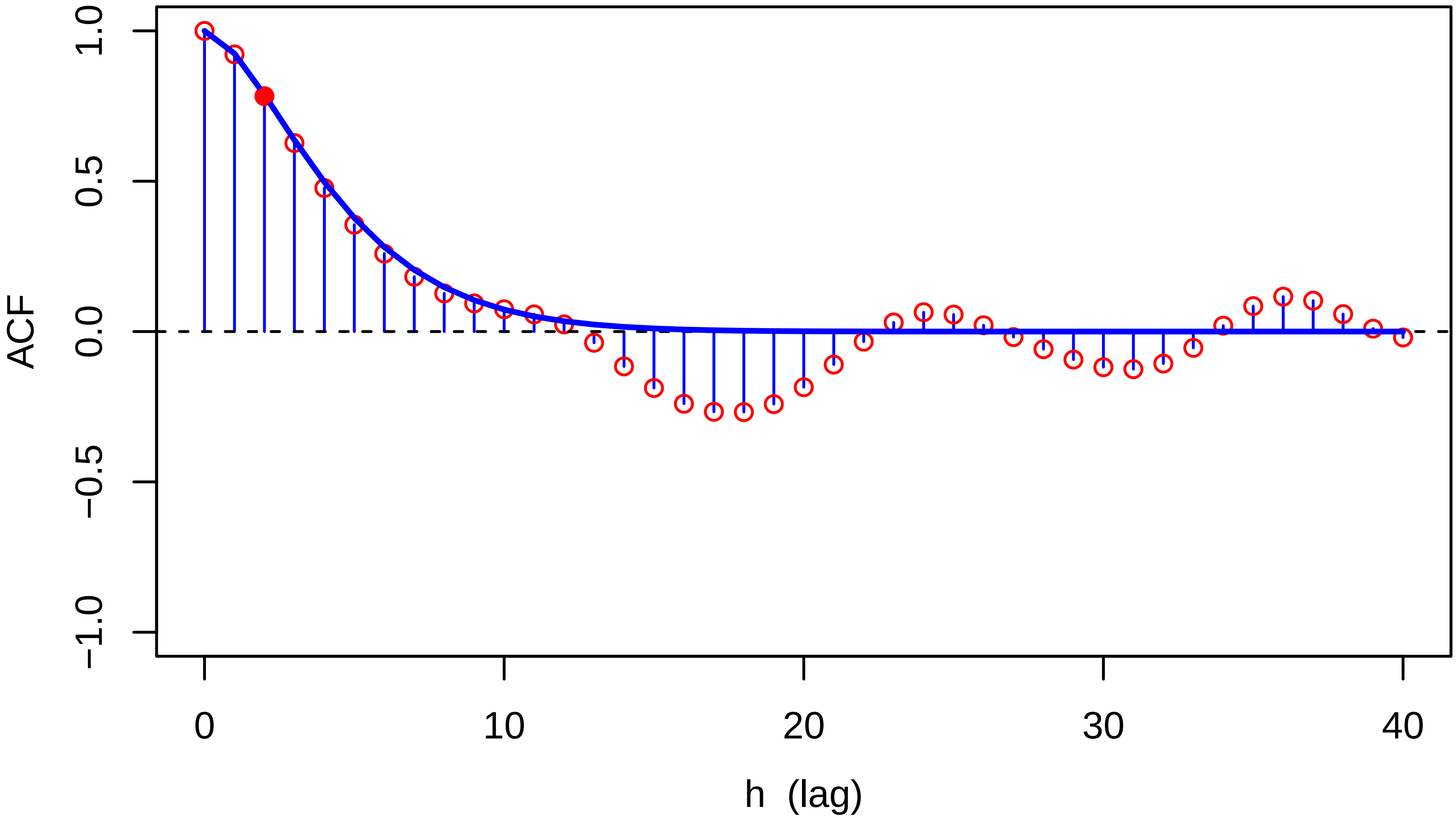
Burg Estimate of PACF for Recruitment Series



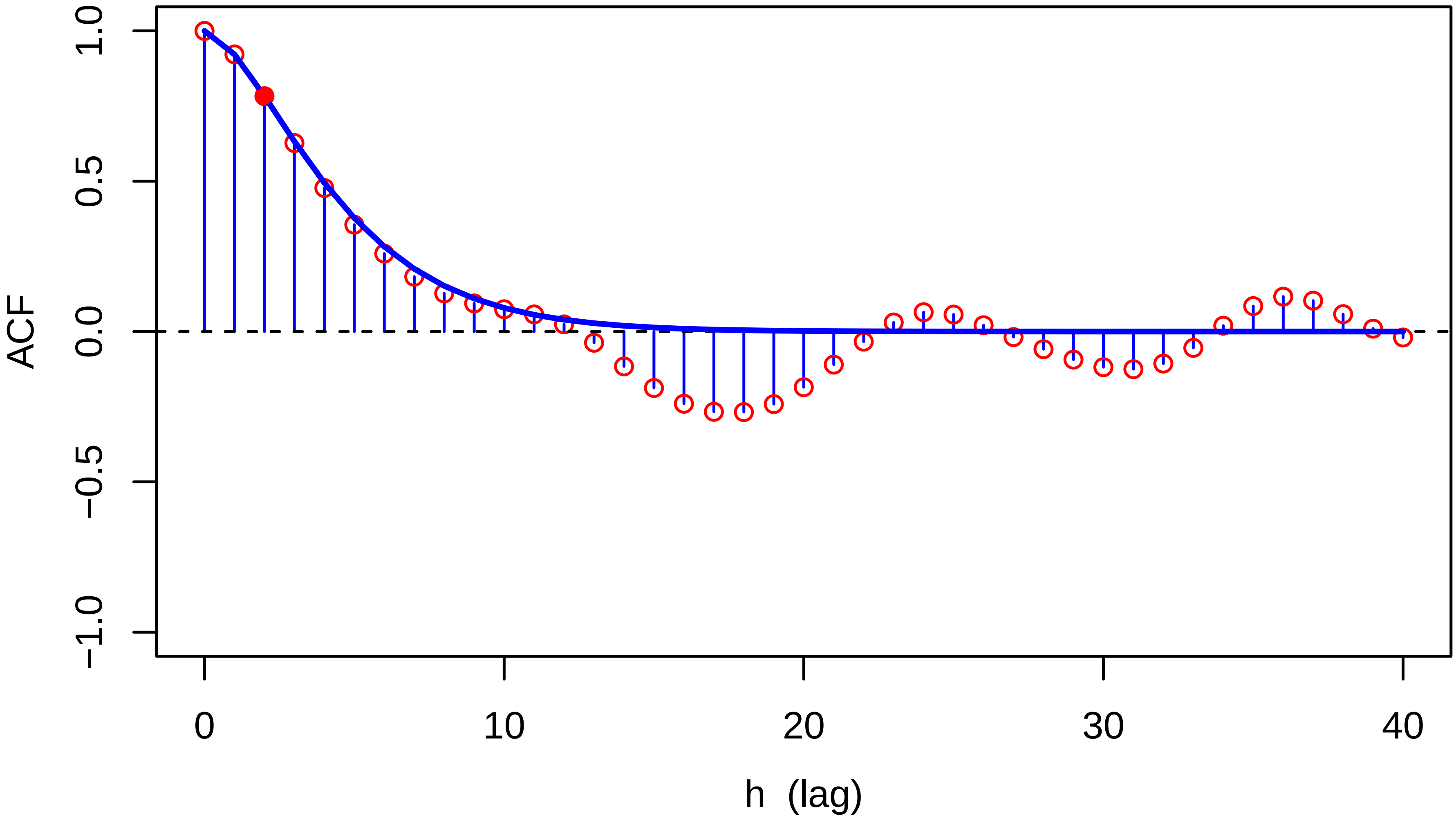
Y–W Estimate of PACF for Recruitment Series



Sample & Burg AR(2) ACFs for Recruitment Series



Sample & Y–W AR(2) ACFs for Recruitment Series



Example – Recruitment Time Series: IV

- for AR(2) model, Burg estimates are

$$\bar{\boldsymbol{\phi}} = \begin{bmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \end{bmatrix} \doteq \begin{bmatrix} 1.3515 \\ -0.4620 \end{bmatrix} \text{ and } \bar{\sigma}^2 \doteq 89.337,$$

as compared to Y–W estimates:

$$\hat{\boldsymbol{\phi}} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} \doteq \begin{bmatrix} 1.3316 \\ -0.4445 \end{bmatrix} \text{ and } \hat{\sigma}^2 \doteq 94.171$$

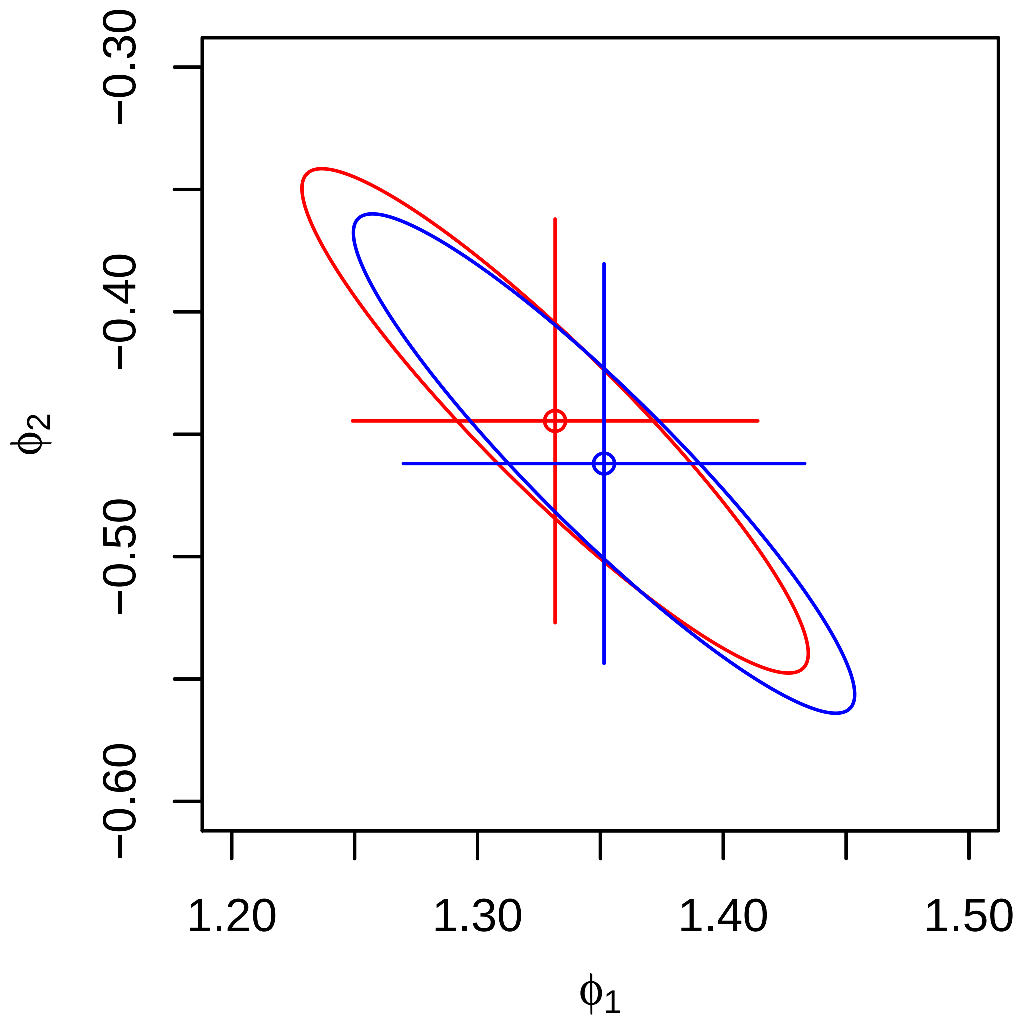
- can determine Burg estimator $\bar{\gamma}(h)$ of ACVF by feeding $\bar{\boldsymbol{\phi}}$ into one of the methods for computing theoretical ARMA ACVFs

- yields

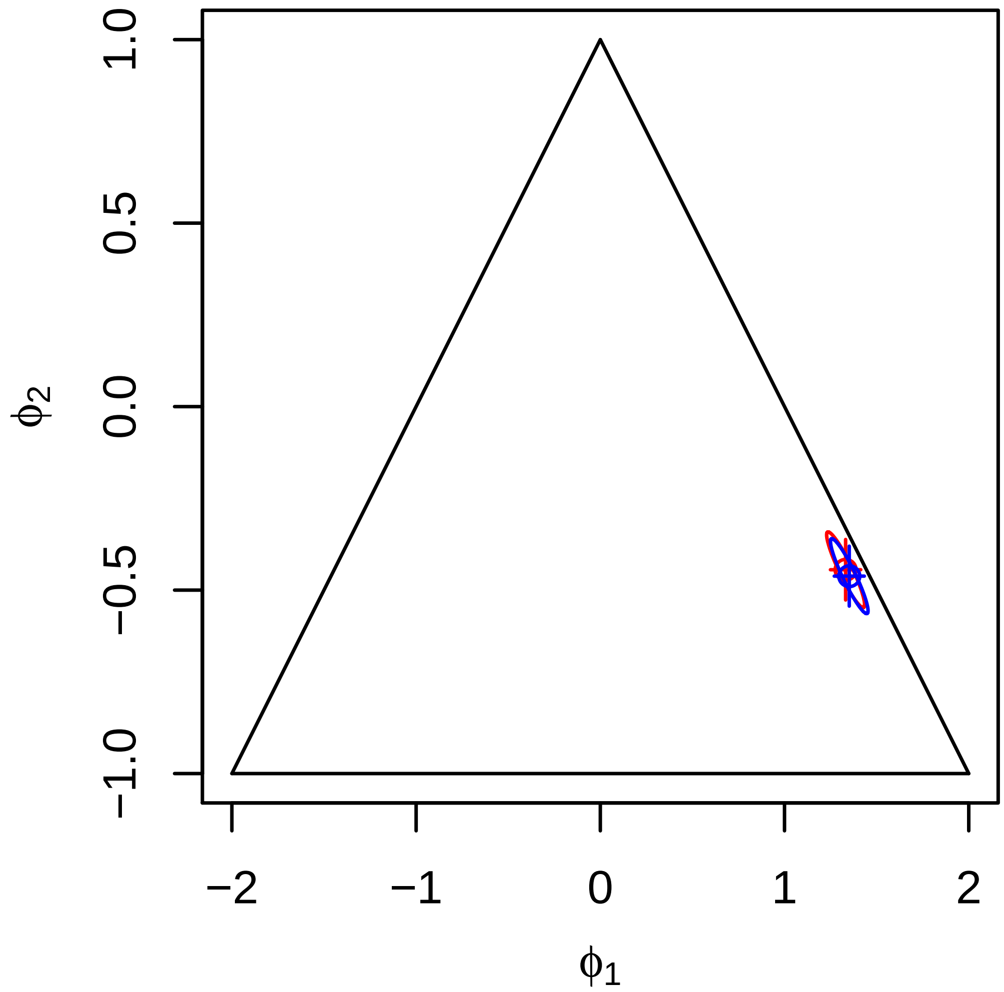
$$\bar{\sigma}^2 \bar{\Gamma}_2^{-1} = \bar{\sigma}^2 \begin{bmatrix} \bar{\gamma}(0) & \bar{\gamma}(1) \\ \bar{\gamma}(1) & \bar{\gamma}(0) \end{bmatrix}^{-1} = \begin{bmatrix} \bar{v}_{1,1} & \bar{v}_{1,2} \\ \bar{v}_{2,1} & \bar{v}_{2,2} \end{bmatrix} \doteq \begin{bmatrix} 0.7866 & -0.7271 \\ -0.7271 & 0.7866 \end{bmatrix}$$

& 95% CIs $[1.2698, 1.4332]$ for ϕ_1 & $[-0.5436, -0.3803]$ for ϕ_2

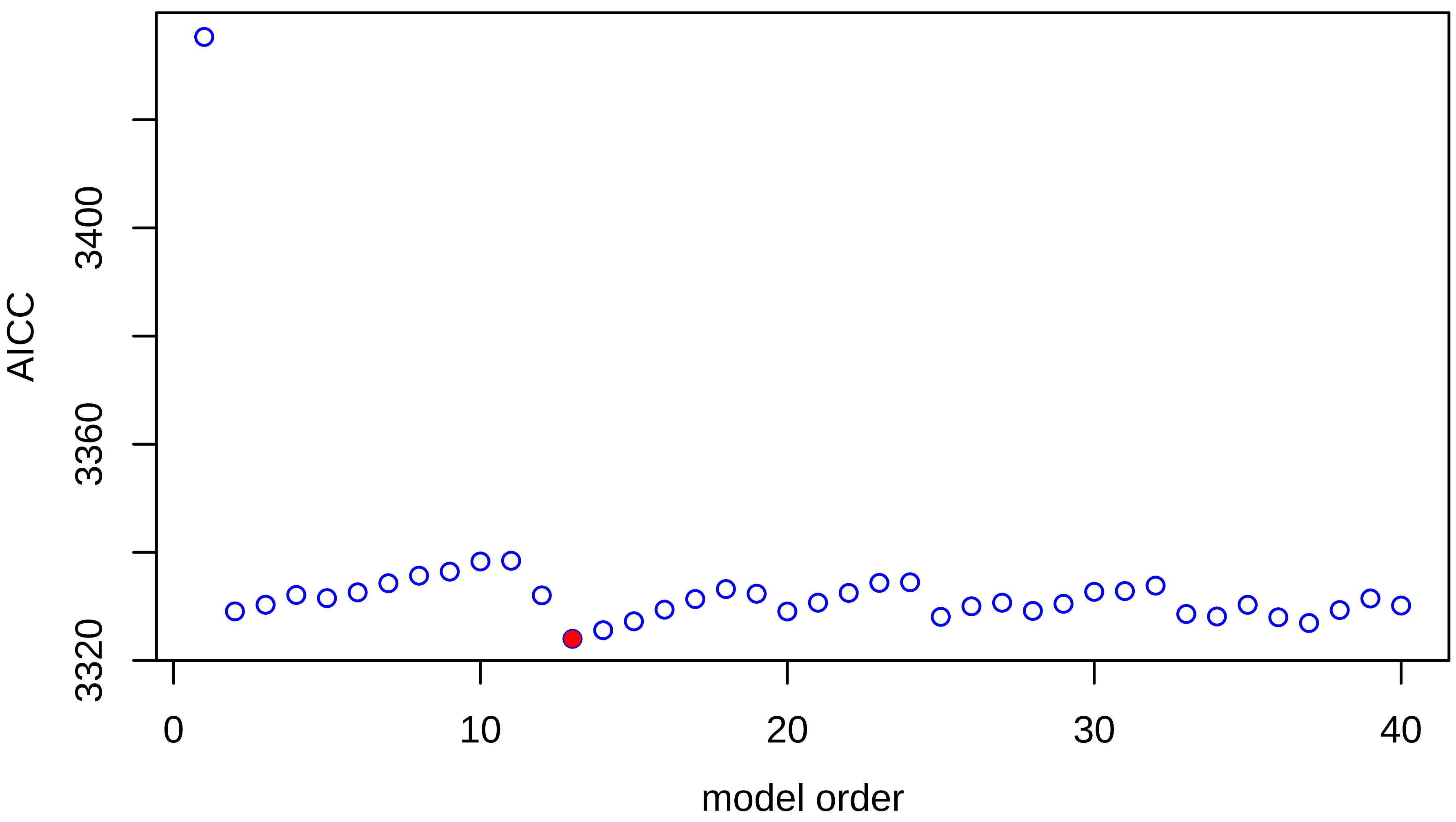
95% Confidence Regions for ϕ (Y–W and Burg)



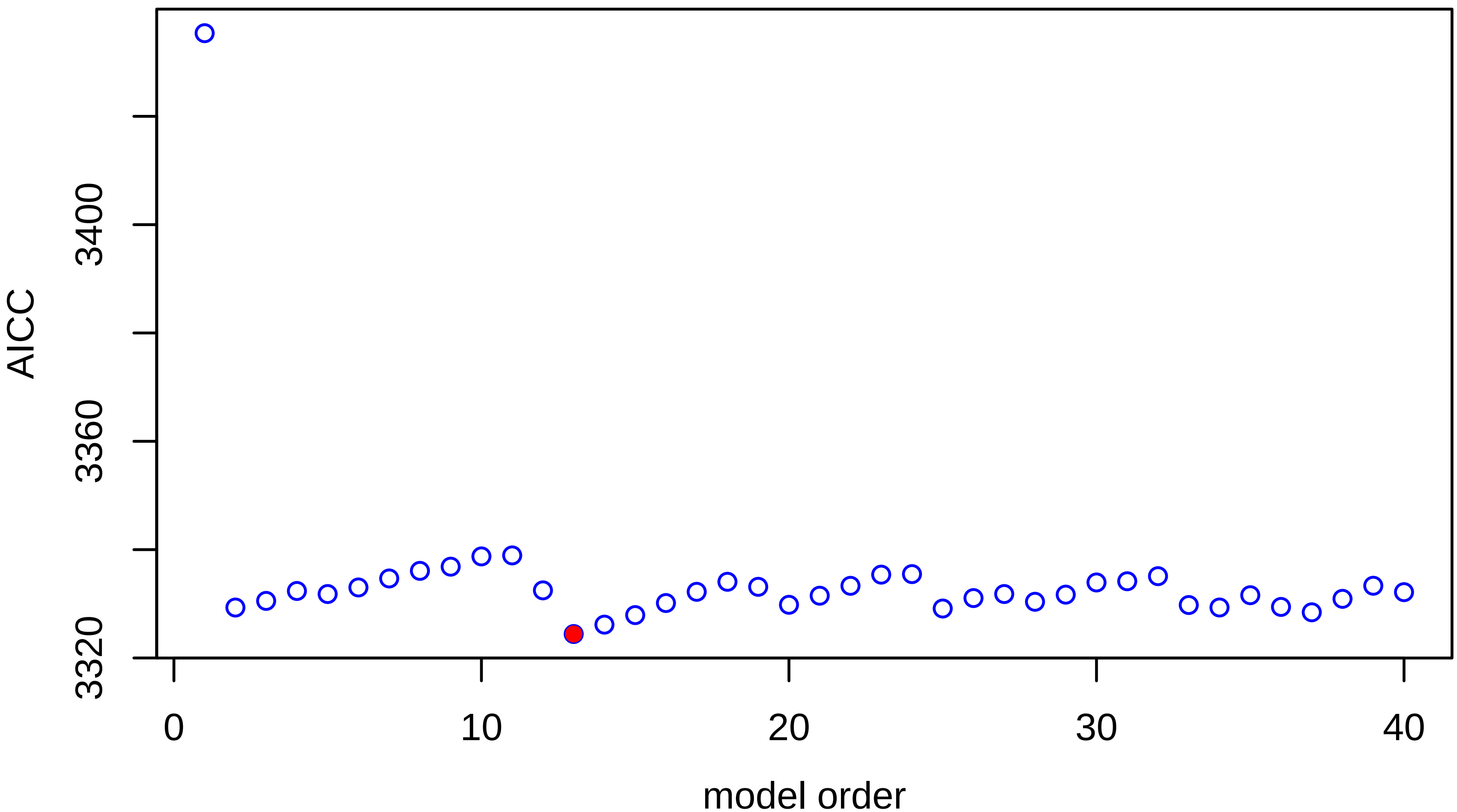
95% Confidence Regions and Causality Region for ϕ



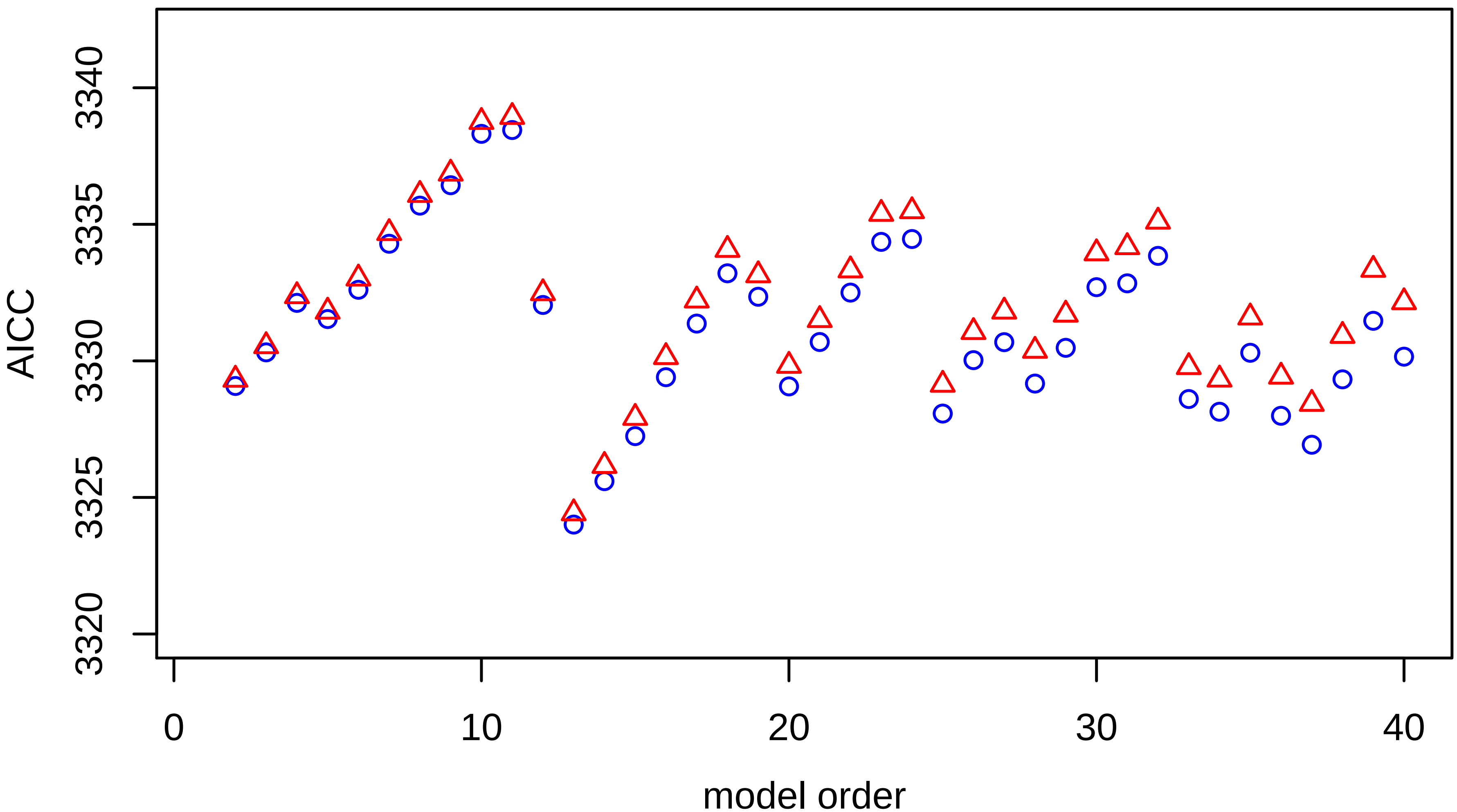
AICC for Recruitment Series (Burg)



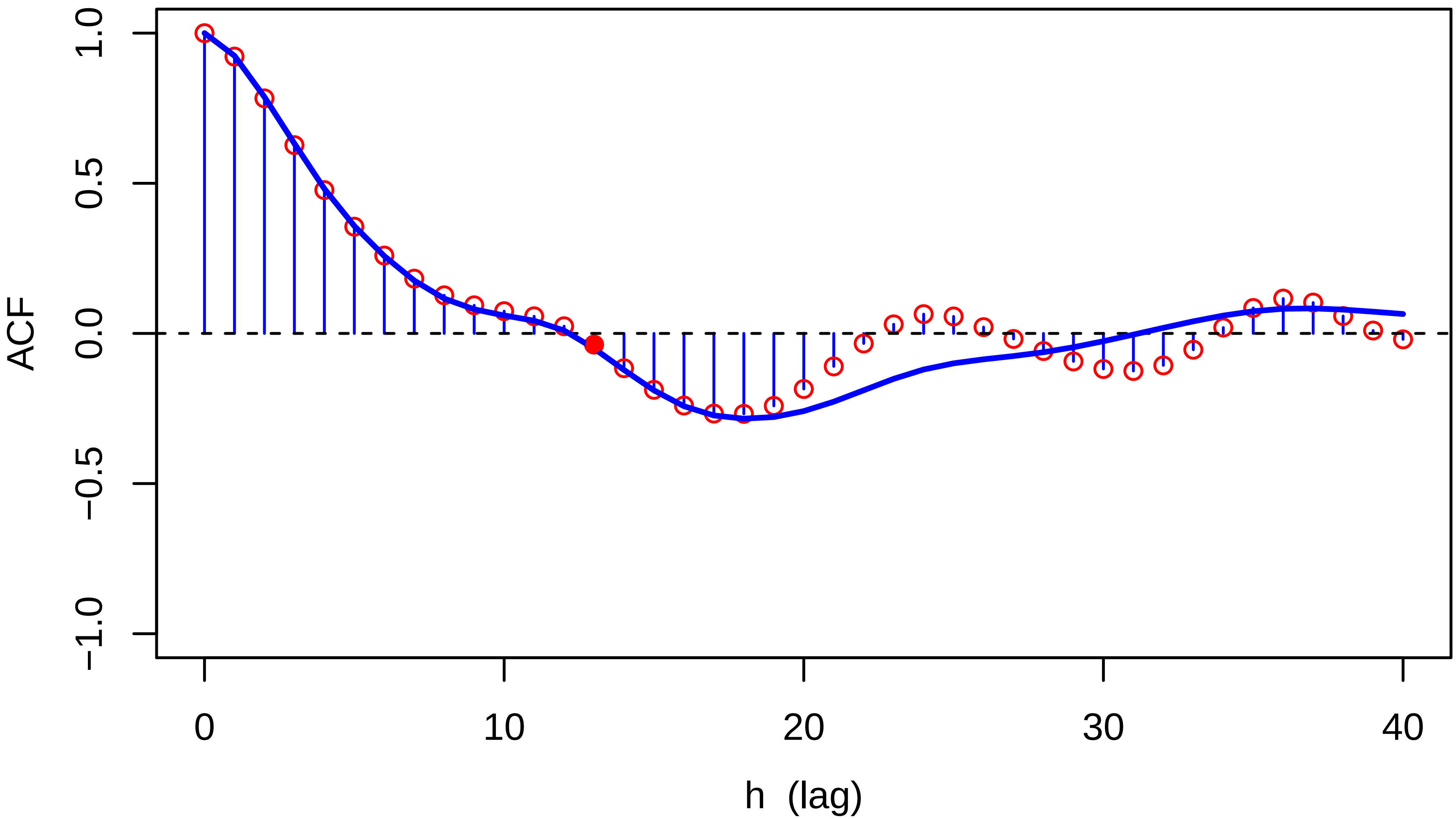
AICC for Recruitment Series (Y–W)

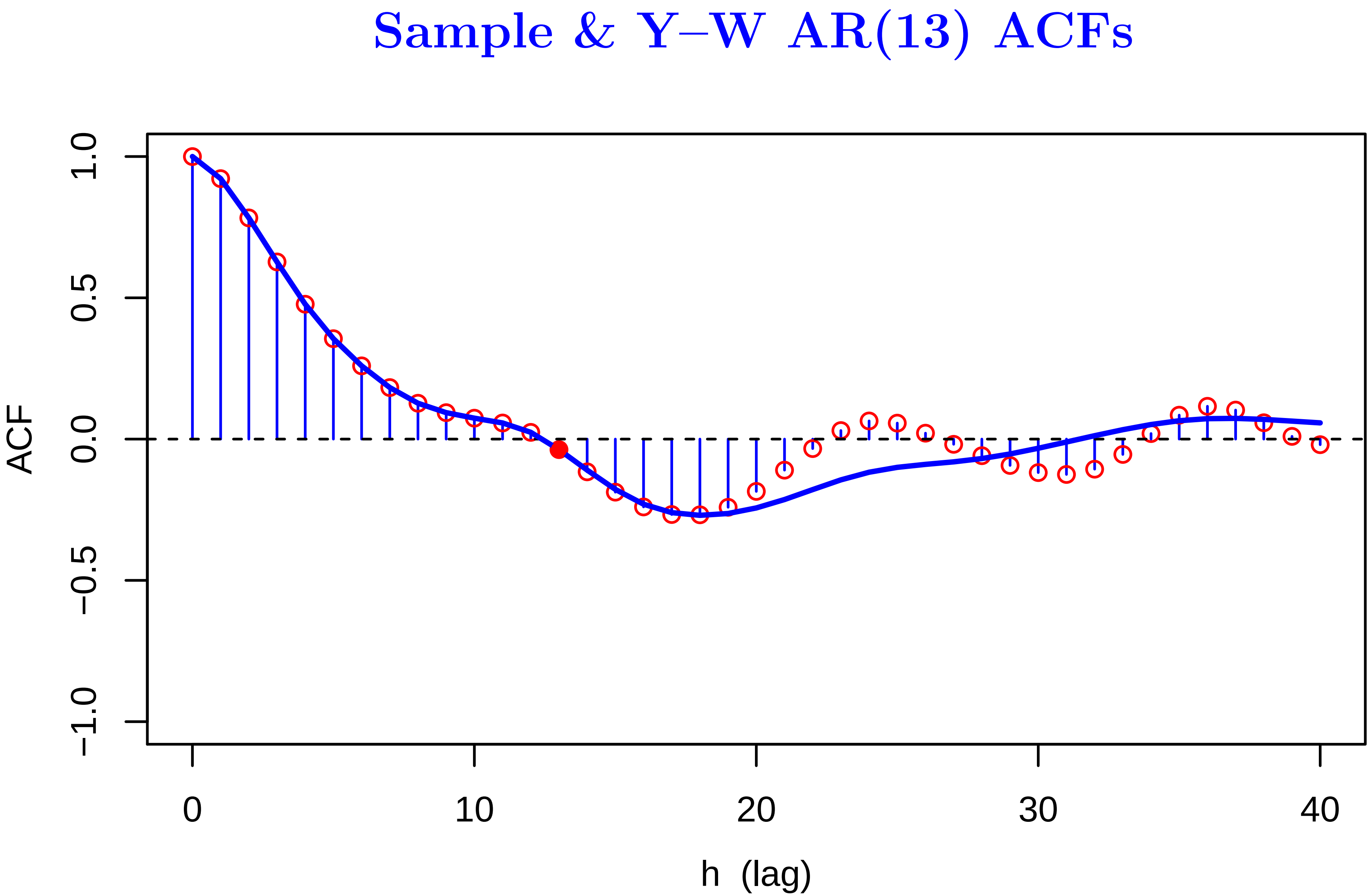


AICC for Recruitment Series (Y–W and Burg)



Sample & Burg AR(13) ACFs




Sample & Y–W AR(13) ACFs
ACF
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30
40
h (lag)

Moment Matching and MA(q) Processes: I

- Y–W & Burg give preliminary estimates of $\boldsymbol{\phi}$ & σ^2 for AR(p)
- Y–W estimator based on moment matching (MM)
- relates $\phi_1, \ldots, \phi_p$ and σ^2 to ACVF values $\gamma(0), \ldots \gamma(p)$ via $p + 1$ linear equations
- solve p equations to get $\boldsymbol{\phi}$, namely,

$$\begin{aligned}
\phi_1\gamma(0) + \phi_2\gamma(1) + \cdots \phi_p\gamma(p-1) &= \gamma(1) \\
\phi_1\gamma(1) + \phi_2\gamma(0) + \cdots \phi_p\gamma(p-2) &= \gamma(2) \\
&\vdots \\
\phi_1\gamma(p-1) + \phi_2\gamma(p-2) + \cdots \phi_p\gamma(0) &= \gamma(p)
\end{aligned}$$

  after which get σ^2 via

$$\sigma^2 = \gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p)$$

- Q: is a similar scheme viable to estimate $\boldsymbol{\theta}$ & σ^2 for MA(q)?

Moment Matching and MA(q) Processes: II

- consider invertible MA(1) model: $X_t = Z_t + \theta Z_{t-1}$ with $|\theta| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- ACVF given by

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

- using $\gamma(0) = \sigma^2(1+\theta^2)$ and $\gamma(1) = \sigma^2\theta$ to express θ in terms of ACVF leads to solving nonlinear equation

$$\frac{\gamma(1)}{\gamma(0)} = \rho(1) = \frac{\theta}{1+\theta^2}, \text{ i.e., need to find roots of } \rho(1)\theta^2 - \theta + \rho(1) = 0$$

Moment Matching and MA(q) Processes: III

- when $\rho(1) \neq 0$, possible solutions are

$$\theta = \frac{1 \pm \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}, \text{ which requires } -\tfrac{1}{2} \le \rho(1) \le \tfrac{1}{2}$$

  for θ to be real-valued (need $-\frac{1}{2} < \rho(1) < \frac{1}{2}$ to satisfy $|\theta| < 1$)
- since $\hat{\rho}(1)$ need *not* obey this constraint, MM can fail to give viable estimators of $\boldsymbol{\theta}$ for MA(q) process
- failure of MM *might* suggest MA(q) model is inappropriate!
- if viable MM estimator $\hat{\theta}$ exists in MA(1) case, estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\hat{\gamma}(0)}{1 + \hat{\theta}^2}$$

BD: 128–129, CC: 150, SS: 115

Innovations Algorithm: I

- alternative to moment matching is innovations algorithm (IA), which involves linear manipulation of $\gamma(h)$'s
- innovations representation for $X_1, X_2, X_3, \ldots,$ is

$$X_{n+1} = \sum_{j=0}^{n} \theta_{n,j} U_{n-j+1}, \quad n = 0, 1, 2, \ldots$$

where $\theta_{n,0} = 1$, other $\theta_{n,j}$'s given by IA, $U_1 = X_1$ and

$$U_{n+1} = X_{n+1} - \widehat{X}_{n+1}, \text{ with } \widehat{X}_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_n, \quad n = 1, 2, \ldots,$$

and $\widehat{X}_1 = 0$ (see overhead XI–19)

- $v_n = \text{var}\,\{U_{n+1}\} = E\{(X_{n+1} - \widehat{X}_{n+1})^2\}$ is associated MSE

Innovations Algorithm: II

- when $\{X_t\}$ is invertible MA(q) process with coefficients $\theta_0 \stackrel{\text{def}}{=} 1$, $\theta_1, \ldots, \theta_q$ & white noise variance σ^2, innovations representation simplifies for all $n \geq q$:

$$X_{n+1} = \sum_{j=0}^{q} \theta_{n,j} U_{n-j+1}, \quad n = q, q+1, q+2, \ldots,$$

  where $\theta_{n,j} \to \theta_j$ and $v_n \to \sigma^2$ as $n \to \infty$

- for comparison, note that MA(q) model says, for all n,

$$X_{n+1} = \sum_{j=0}^{q} \theta_j Z_{n-j+1}$$

- with infinite past rather than just $X_1, \ldots, X_n$, can argue that $\theta_{n,j}$ & U_{n-j+1} would morph into θ_j & Z_{n-j+1}

Innovations Algorithm: III

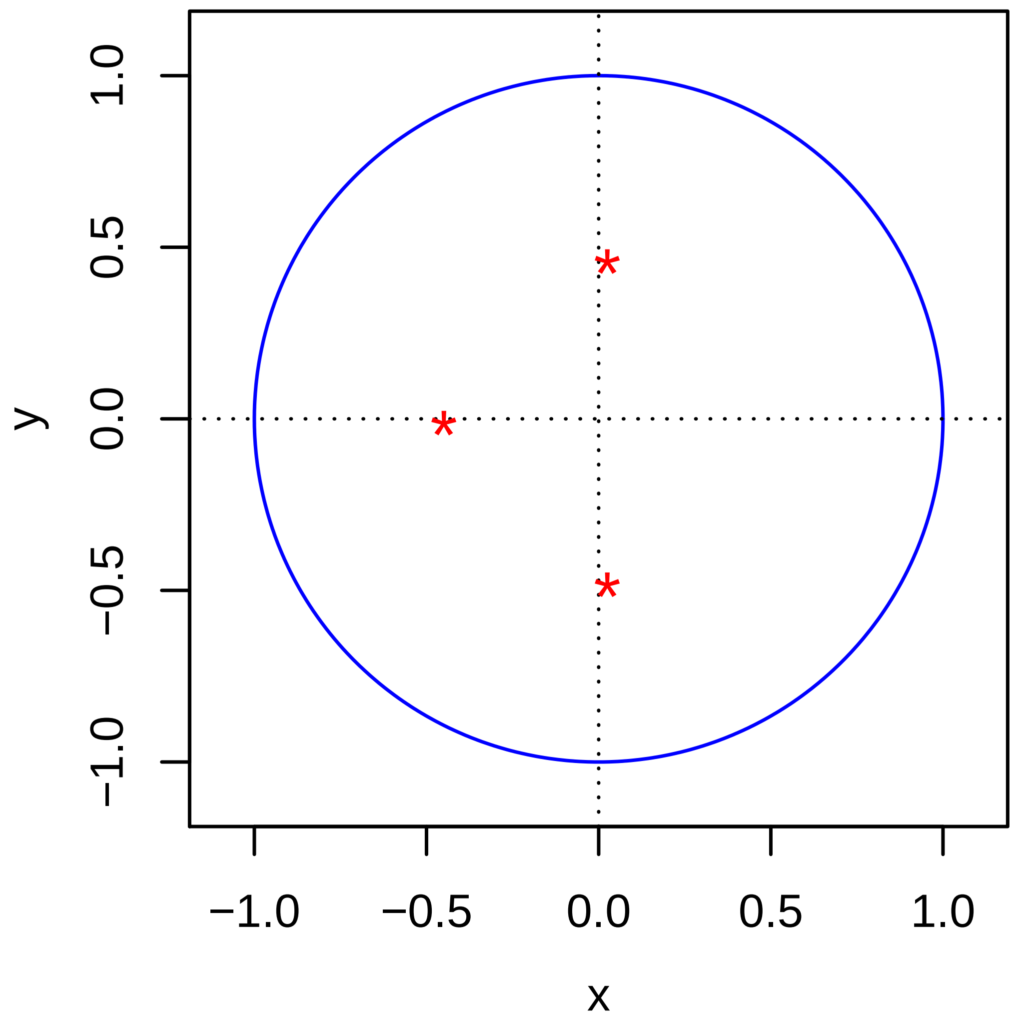
- convergence of $\theta_{n,j}$ to θ_j and v_n to σ^2 can be rapid or painfully slow, depending on how close roots of $\theta(z)$ are to unit circle
- as examples, let's consider three MA(3) processes:

$$X_t = Z_t + 0.4Z_{t-1} + 0.2Z_{t-2} + 0.1Z_{t-3} \qquad (1)$$

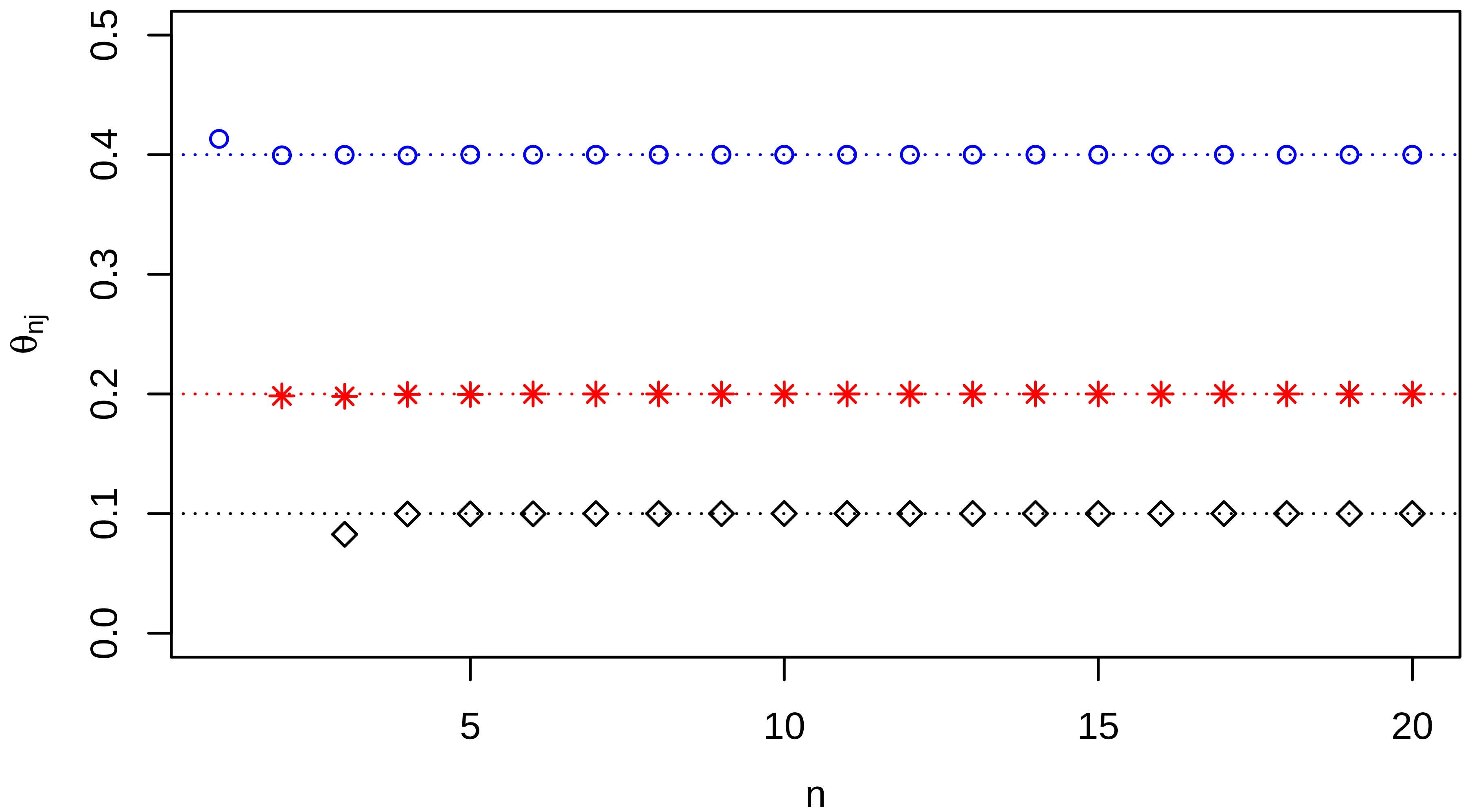
$$X_t = Z_t + 0.8Z_{t-1} + 0.8Z_{t-2} + 0.8Z_{t-3} \qquad (2)$$

$$X_t = Z_t + 0.84Z_{t-1} + 0.88Z_{t-2} + 0.92Z_{t-3} \qquad (3)$$

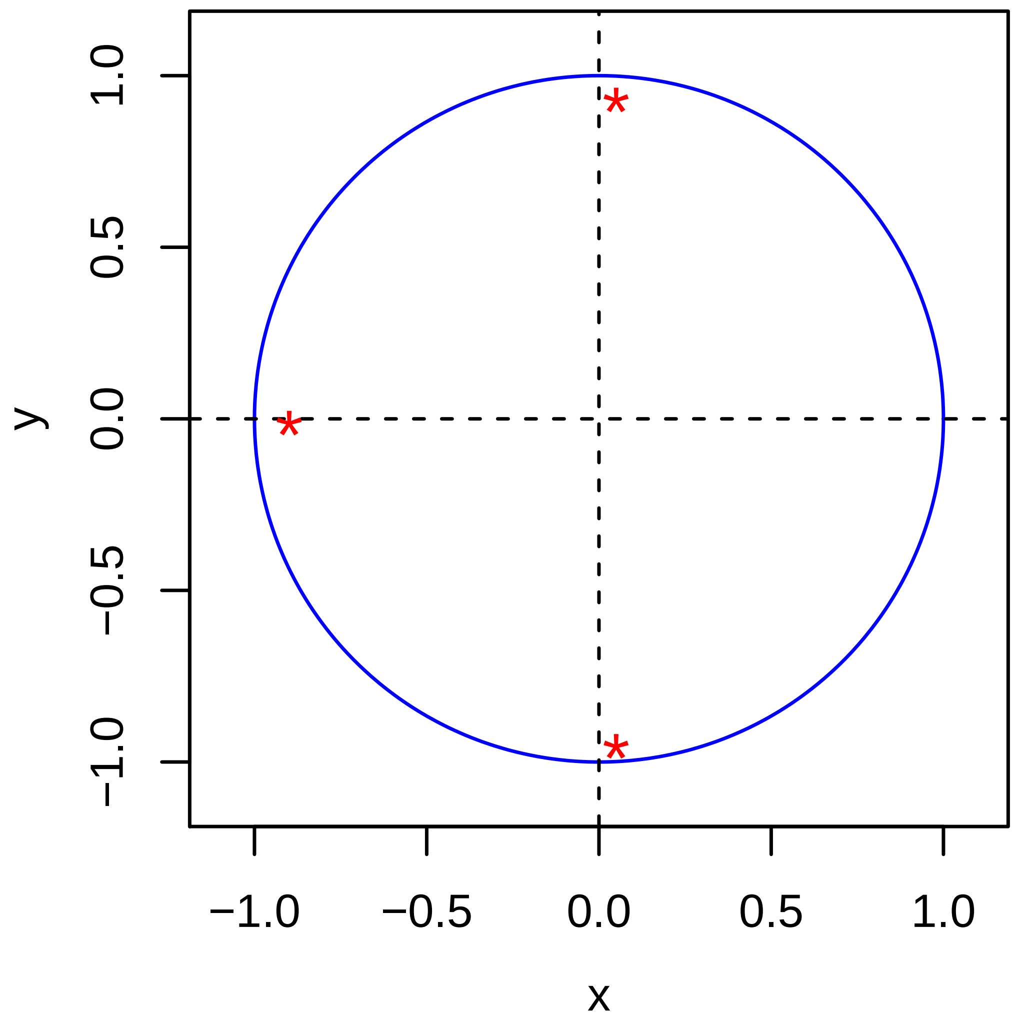
Reciprocal Roots Plot for First MA(3) Process



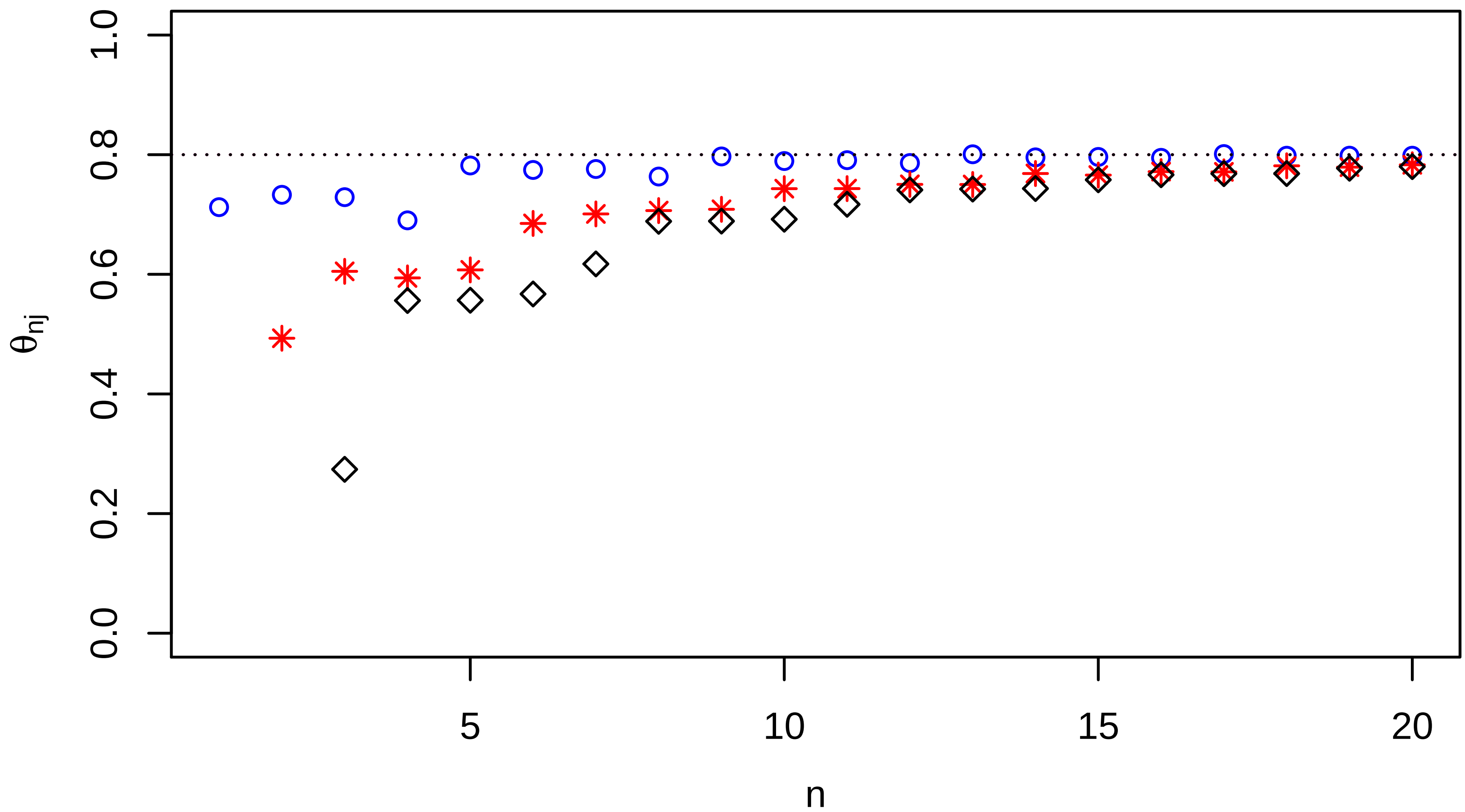
Convergence of $\theta_{n,j}$'s to θ_j for First MA(3) Process



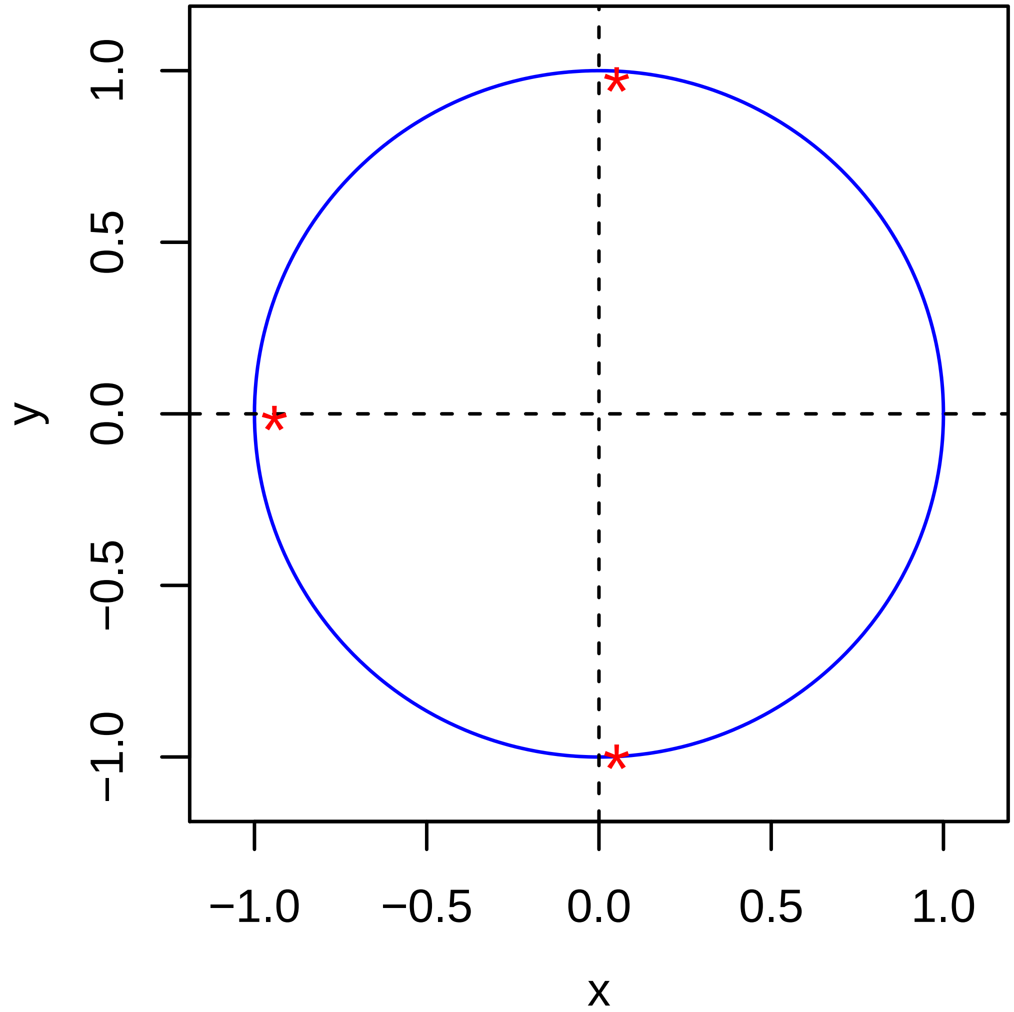
Reciprocal Roots Plot for Second MA(3) Process



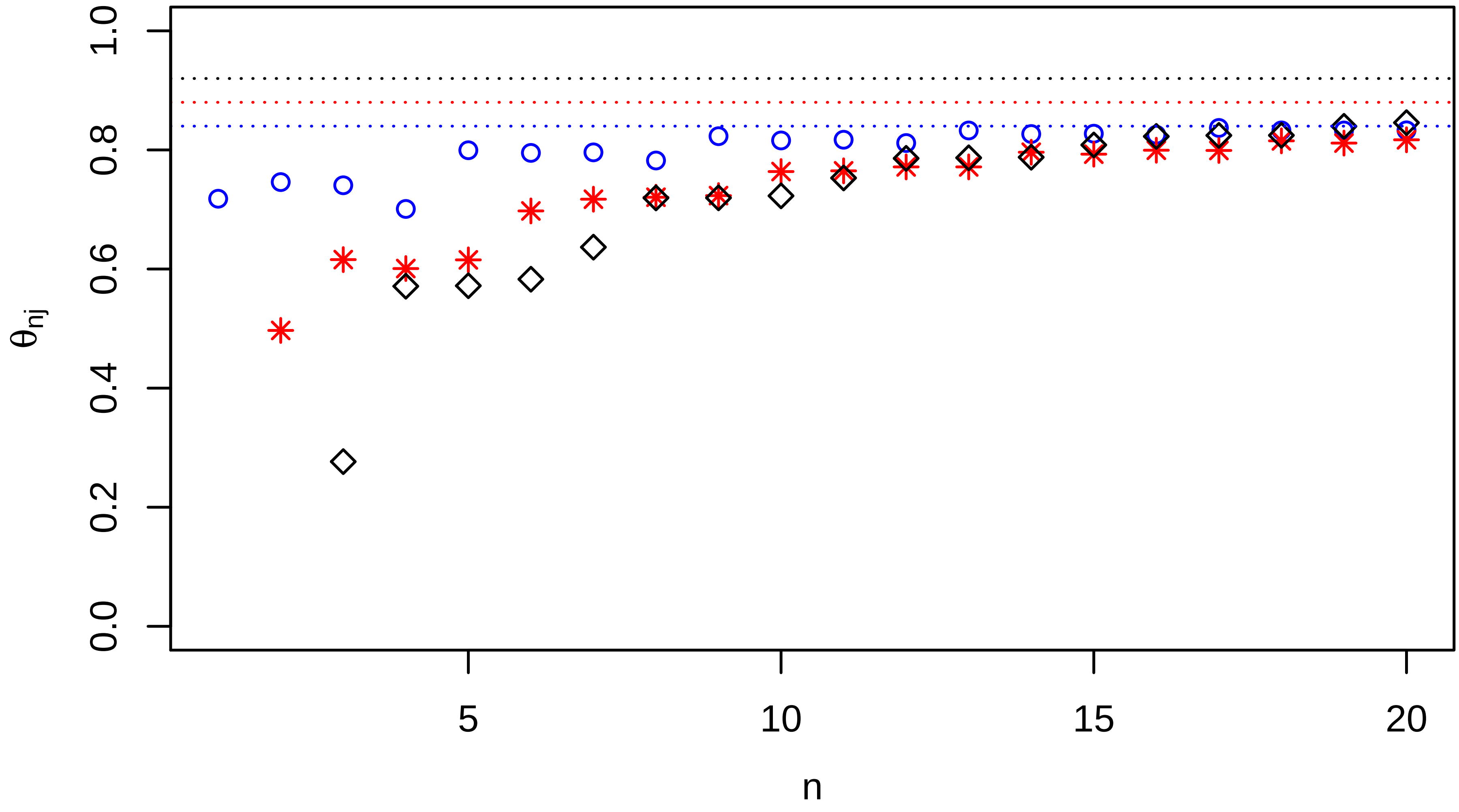
Convergence of $\theta_{n,j}$'s to θ_j for Second MA(3) Process



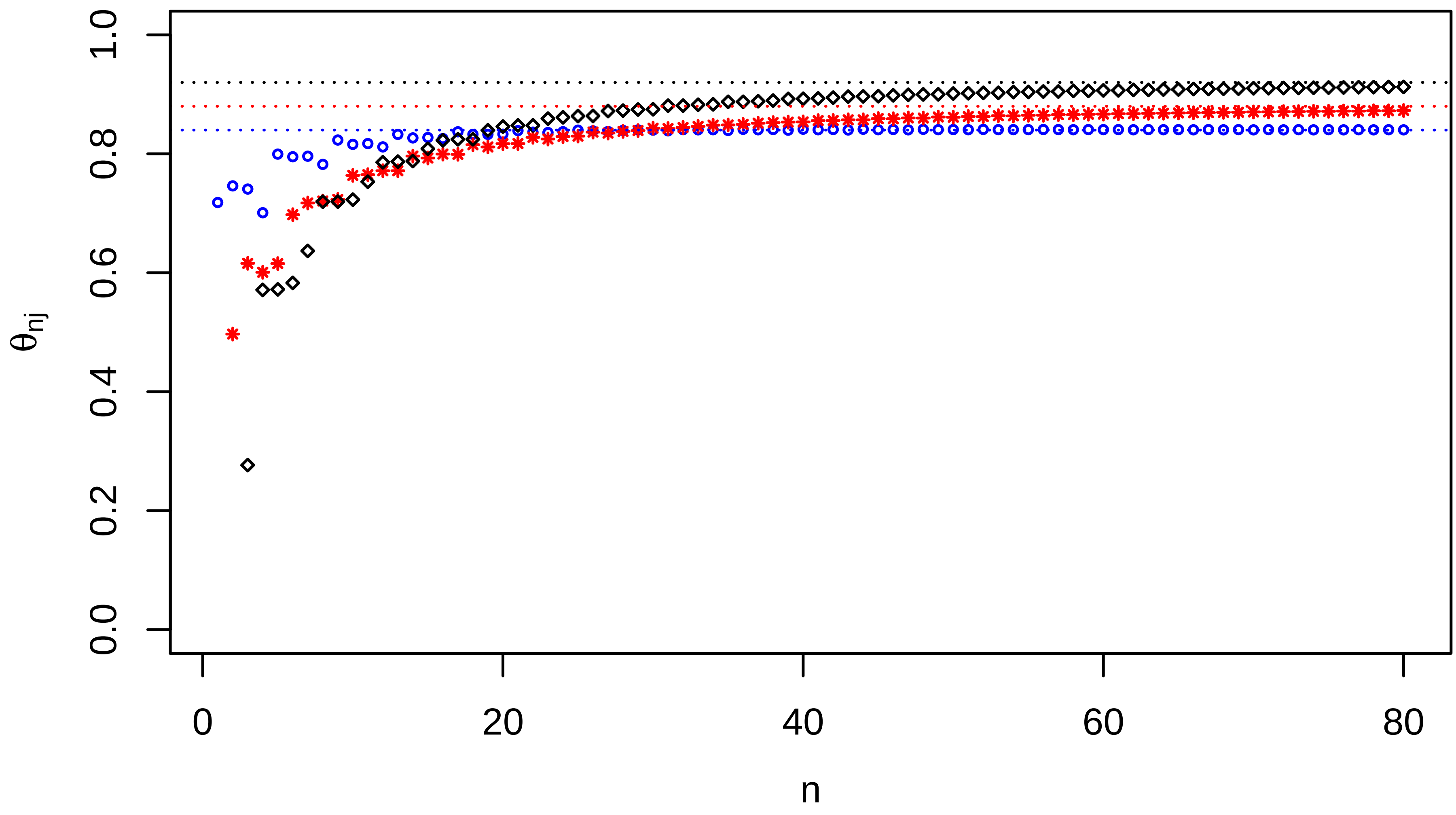
Reciprocal Roots Plot for Third MA(3) Process



Convergence of $\theta_{n,j}$'s to θ_j for Third MA(3) Process



Convergence of $\theta_{n,j}$'s to θ_j for Third MA(3) Process



Innovations Algorithm: IV

- IA munches on $\gamma(h)$'s and spits out $\theta_{n,j}$'s and v_n's
- IA estimators of $\boldsymbol{\theta}$ and σ^2 for MA(q) process gotten by letting IA munch on $\hat{\gamma}(h)$'s instead
- let $\hat{\theta}_{n,j}$ and $\hat{v}_n$ denote what IA gives when handed $\hat{\gamma}(h)$'s
- IA estimators $\hat{\boldsymbol{\theta}}$ & $\hat{\sigma}^2$ of $\boldsymbol{\theta}$ & σ^2 given by $\hat{\theta}_j = \hat{\theta}_{n,j}$ & $\hat{\sigma}^2 = \hat{v}_n$
- large sample theory for $\hat{\boldsymbol{\theta}}$ & $\hat{\sigma}^2$ more complicated than that for Y–W and Burg (see B&D, p. 133)

Innovations Algorithm: V

- theory says: for invertible MA(q) process satisfying certain regularity conditions and for any fixed $k > 0$, normalized vector

$$n^{1/2}\left[\hat{\theta}_{m,1} - \theta_1, \ldots, \hat{\theta}_{m,k} - \theta_k\right]'$$

converges to multivariate normal with mean vector $\mathbf{0}$ and covariance matrix A whose (i, j)th element is

$$A_{i,j} = \sum_{l=1}^{\min\{i,j\}} \theta_{i-l}\theta_{j-l}; \text{ in particular, } \operatorname{var}\{\hat{\theta}_{m,i}\} \approx \frac{A_{i,i}}{n} = \frac{1}{n}\sum_{l=0}^{i-1}\theta_l^2$$

(here $\theta_0 \stackrel{\text{def}}{=} 1$ and $\theta_l \stackrel{\text{def}}{=} 0$ for $l > q$)

- need $m \ll n$, but m must be large enough so $E\{\hat{\theta}_{m,i}\} \approx \theta_i$
- thus as n grows, m needs to grow also, but at a slower rate

Confidence Intervals and Regions for $\boldsymbol{\theta}$

- can use large sample distribution to get approximate confidence intervals for individual θ_j's or confidence region for vector $\boldsymbol{\theta}$
- with $\hat{\theta}_0 \stackrel{\text{def}}{=} 1$, approx. 95% confidence interval for θ_j given by

$$\left[\hat{\theta}_j - 1.96\left(\frac{\sum_{l=0}^{j-1}\hat{\theta}_l^2}{n}\right)^{1/2}, \hat{\theta}_j + 1.96\left(\frac{\sum_{l=0}^{j-1}\hat{\theta}_l^2}{n}\right)^{1/2}\right]$$

- approx. 95% confidence region for $\boldsymbol{\theta}$ is set of all $\boldsymbol{\theta}$'s such that

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{A}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \le \frac{\chi^2_{0.95}(q)}{n},$$

where $q \times q$ matrix $\hat{A}$ has (i, j)th element given by

$$\hat{A}_{i,j} = \sum_{l=1}^{\min\{i,j\}} \hat{\theta}_{i-l}\hat{\theta}_{j-l}$$

Order Selection for MA(q) Processes

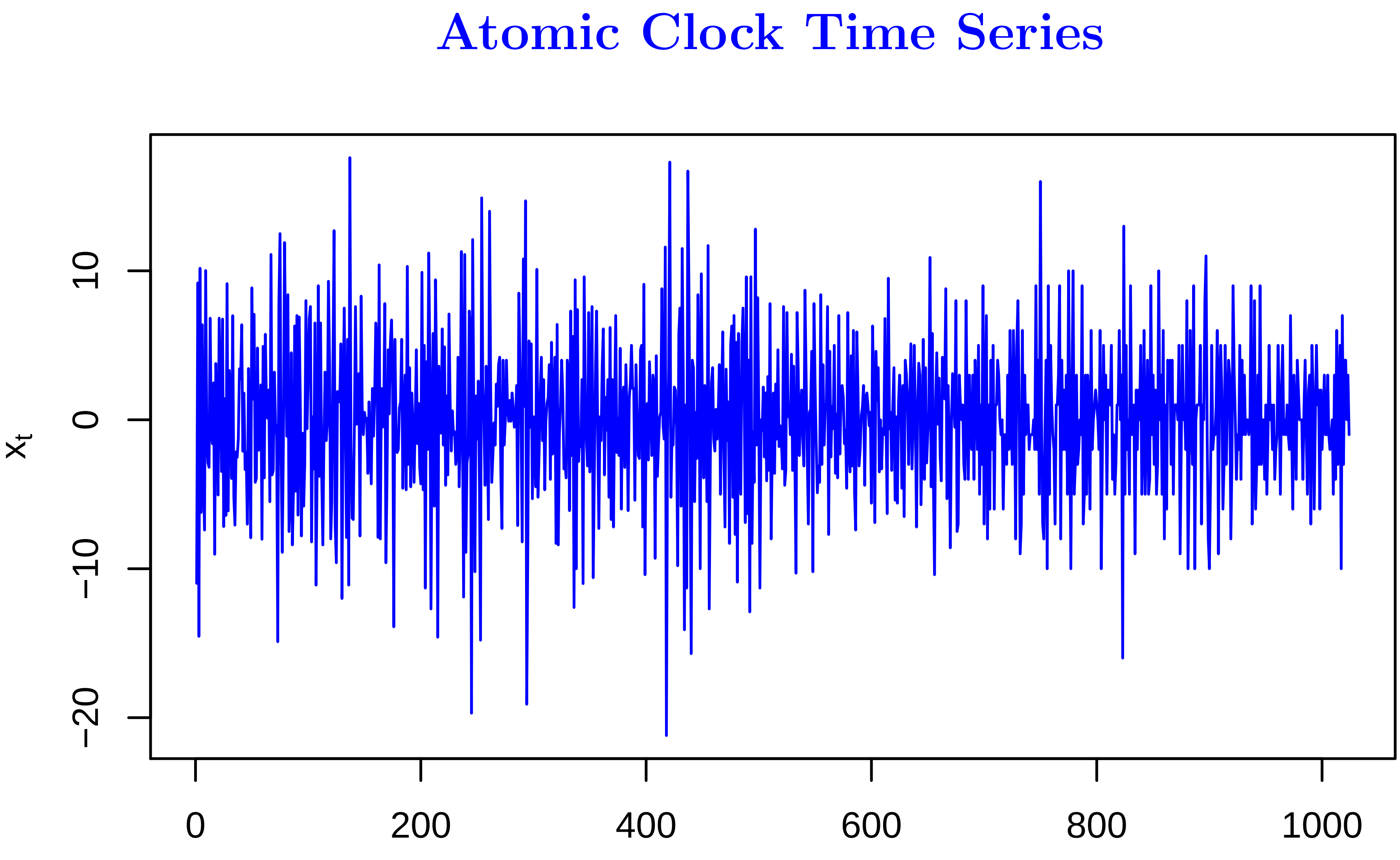
1. for large n, sample ACF RVs for MA(q) time series at lags $h > q$ are approximately $\mathcal{N}(0, w_{h,h}/n)$, where, based upon Bartlett's formula,

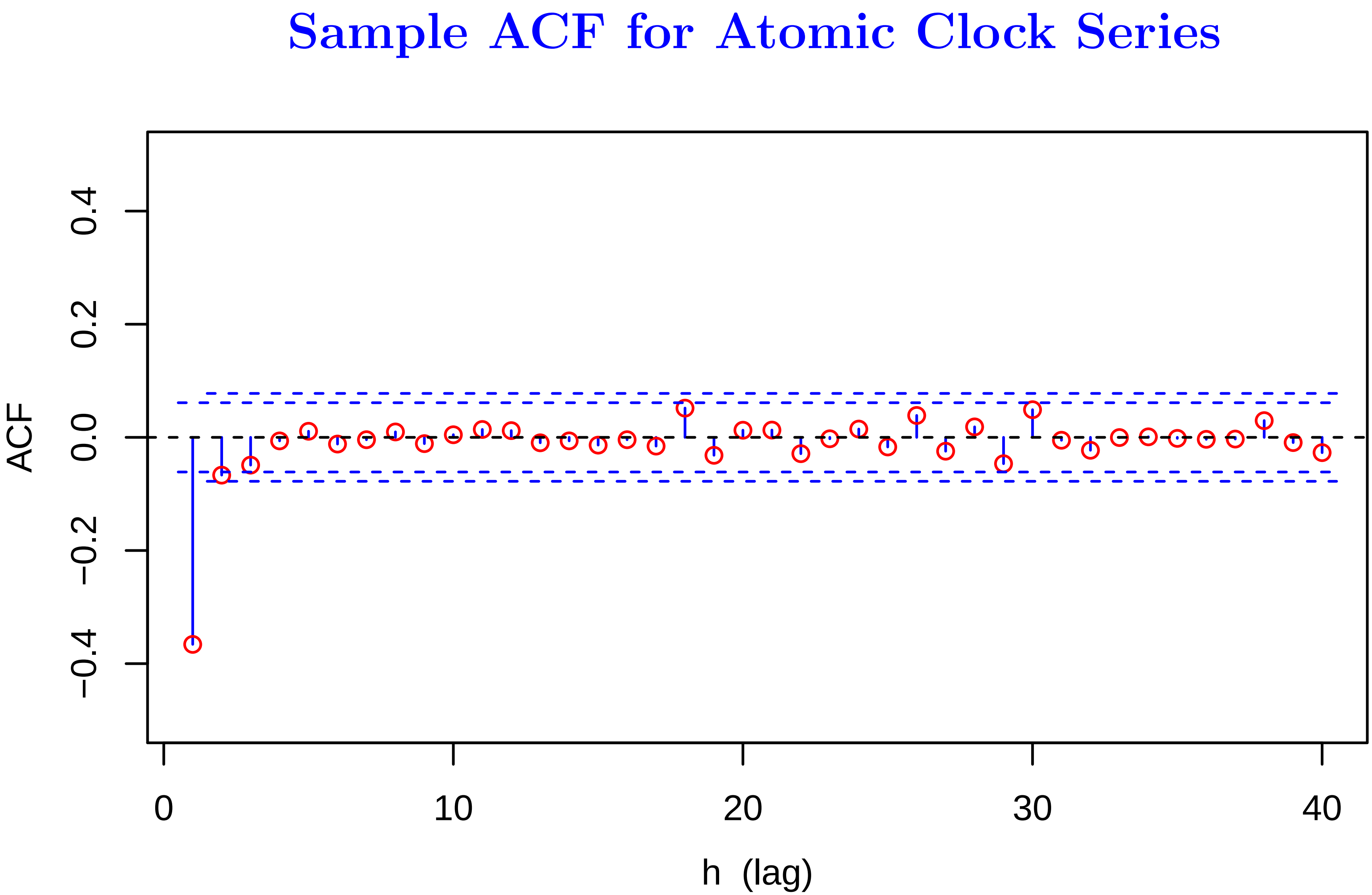
$$w_{h,h} = 1 + 2\rho^2(1) + \cdots + 2\rho^2(q)$$

($\hat{\rho}(h)$ should fall between $\pm 1.96(w_{h,h}/n)^{1/2}$ with prob. $\approx 95\%$)

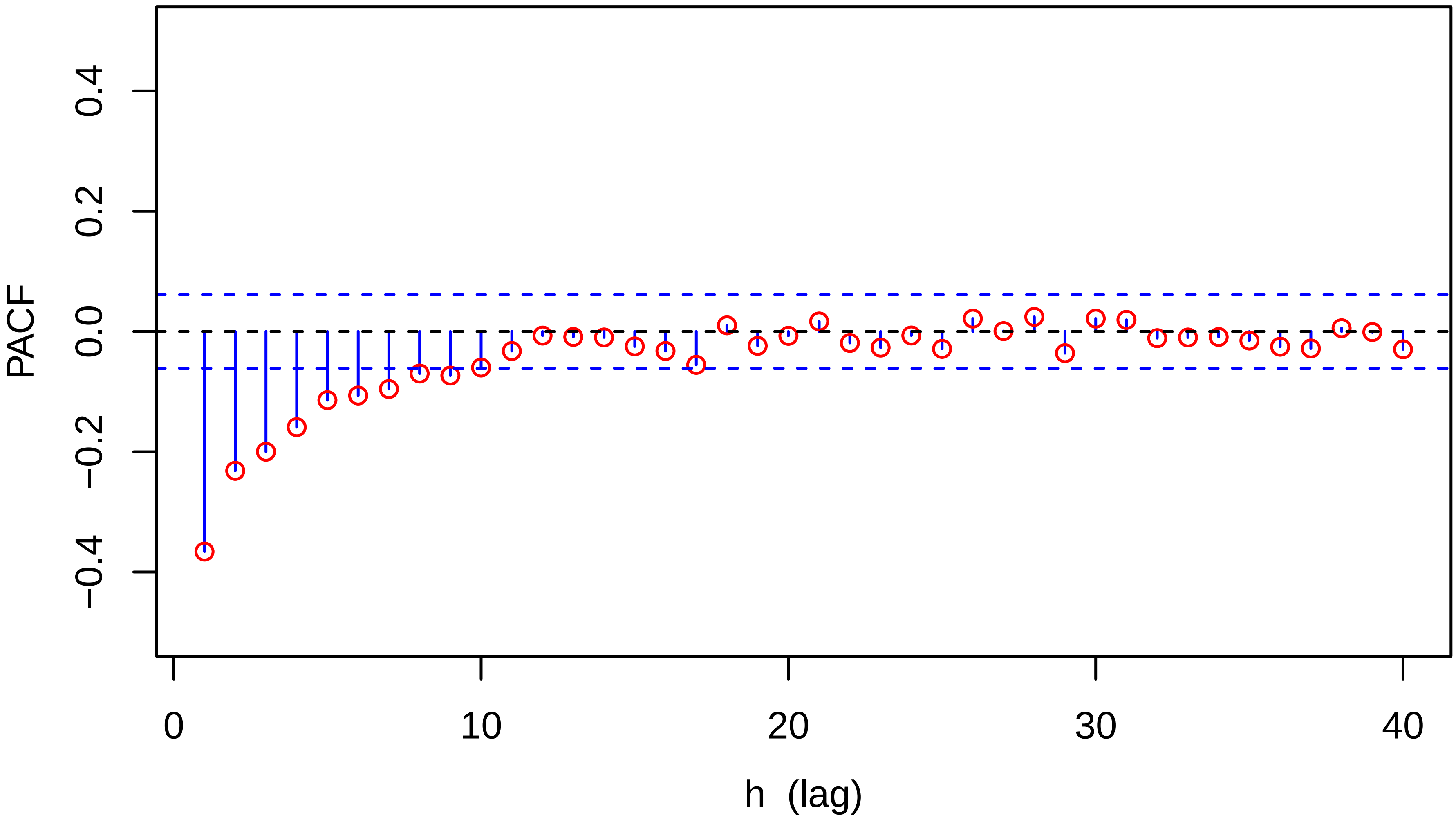
2. in a similar manner, can base order selection on IA estimator $\hat{\boldsymbol{\theta}}$, using its large sample theory to assess variability

3. can also select q to be minimizer of AICC statistic, with likelihood for each order being evaluated using IA estimator $\hat{\boldsymbol{\theta}}$

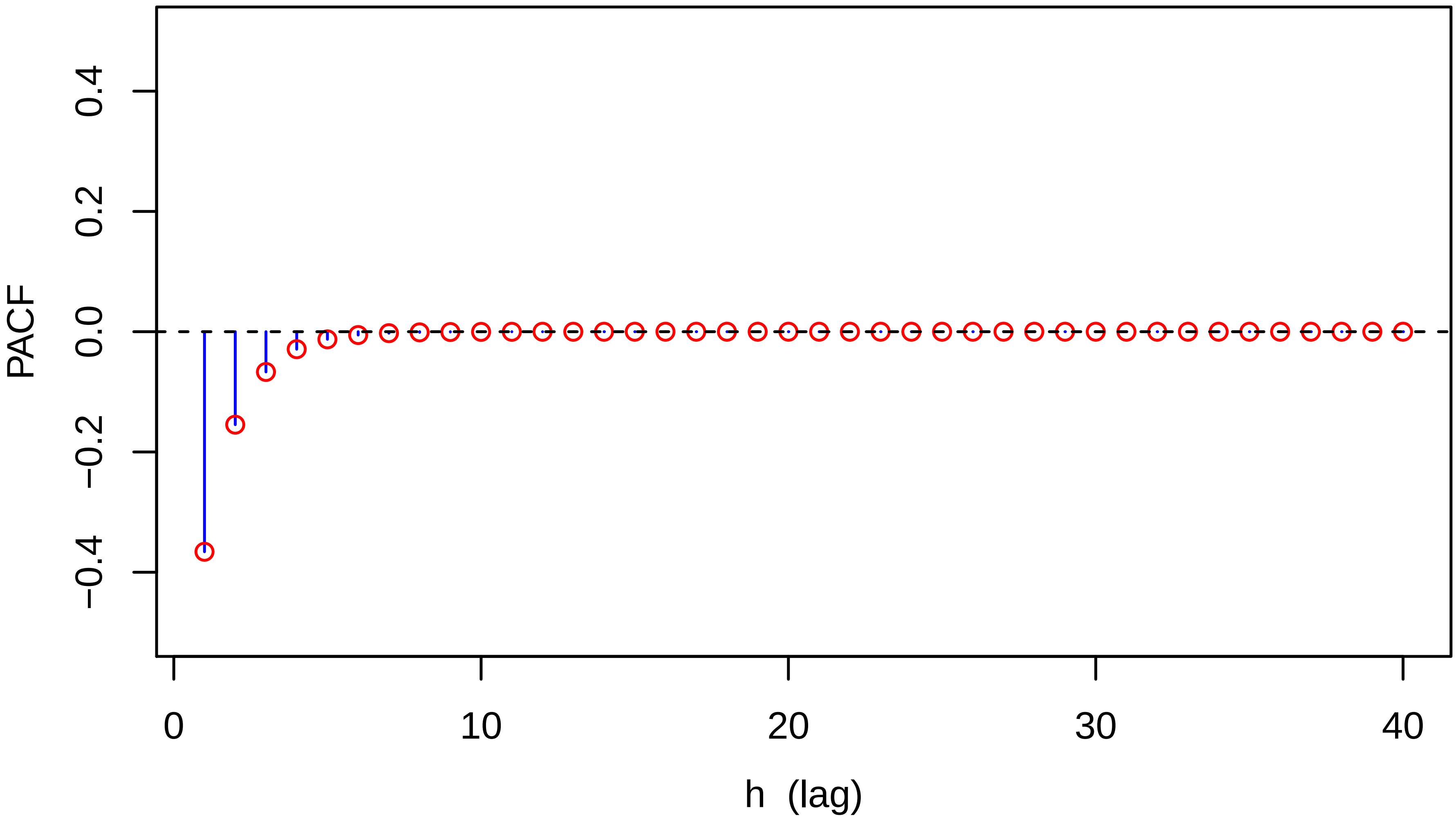

Atomic Clock Time Series
10
0
−10
−20
x_t
0
200
400
600
800
1000
t


Sample ACF for Atomic Clock Series
ACF
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
h (lag)

Sample PACF for Atomic Clock Series



PACF for MA(1) Process with $\theta \doteq -0.4352$



Moment Matching for Atomic Clock Series

- assuming $q = 1$, can do moment matching since $\hat{\rho}(1) \doteq -0.3659$ and hence $|\hat{\rho}(1)| < 0.5$, as required
- possible estimates are

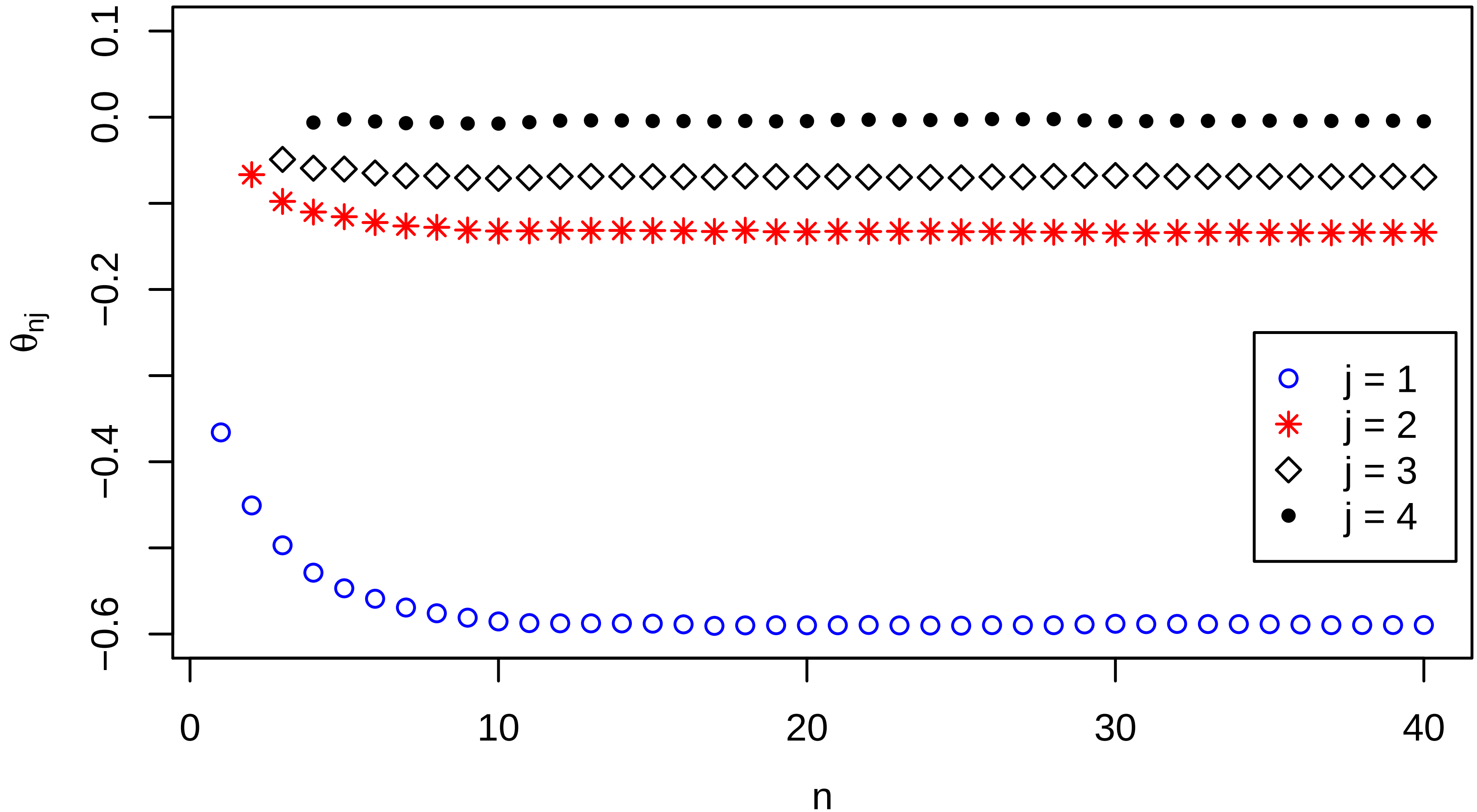
$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4\hat{\rho}^2(1)}}{2\hat{\rho}(1)}; \text{ i.e., either } \hat{\theta} \doteq -2.2978 \text{ or } \hat{\theta} \doteq -0.4352$$

- two estimates are reciprocals of one another
- choice $\hat{\theta} \doteq -0.4352$ corresponds to invertible MA(1) model
- corresponding estimate of σ^2 is

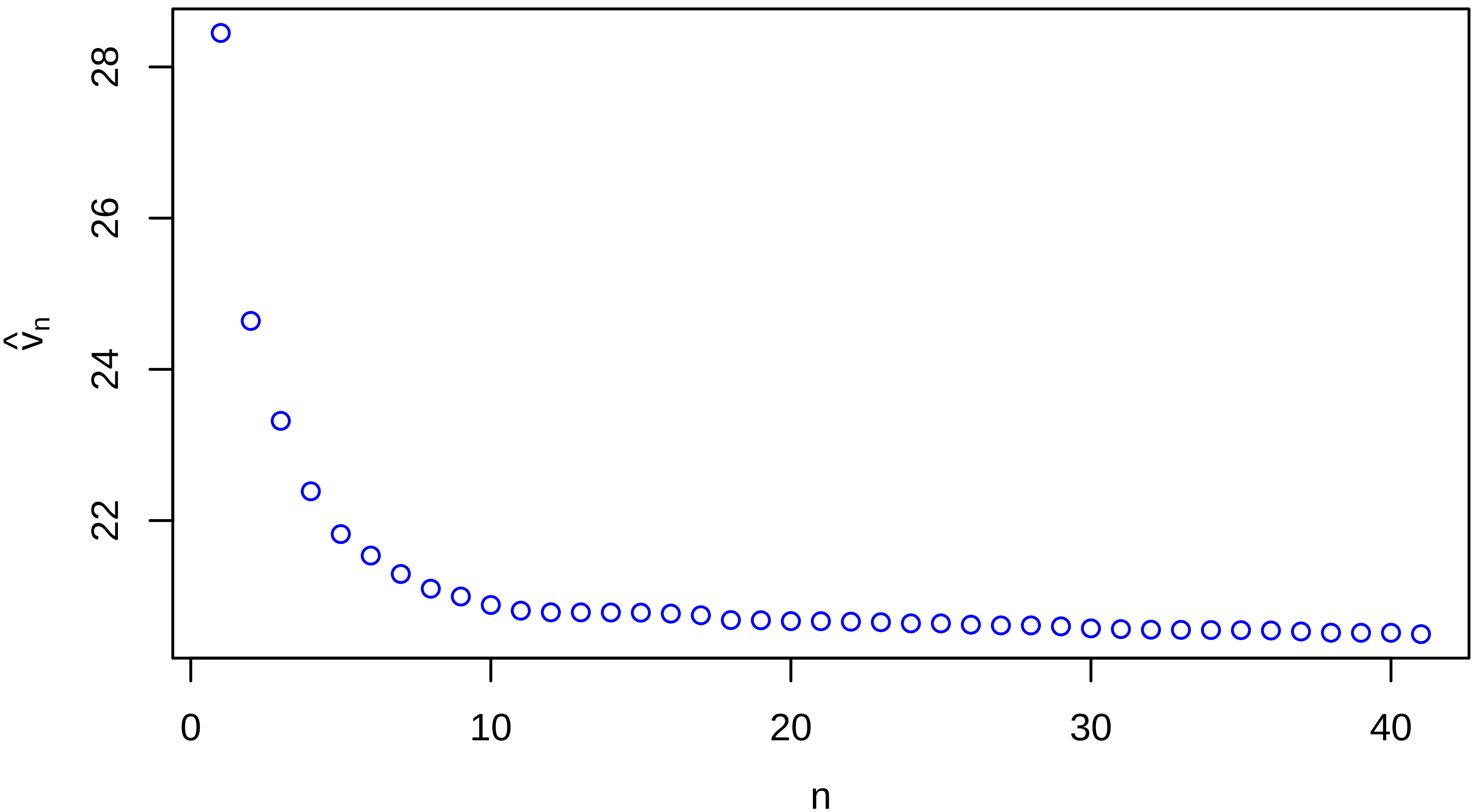
$$\hat{\sigma}^2 = \frac{\hat{\gamma}(0)}{1 + \hat{\theta}^2} \doteq 23.919$$

- assuming $q = 2$, ... hmmm ...

Convergence of $\hat{\theta}_{n,j}$'s for Atomic Clock Series



Convergence of $\hat{v}_n$'s for Atomic Clock Series



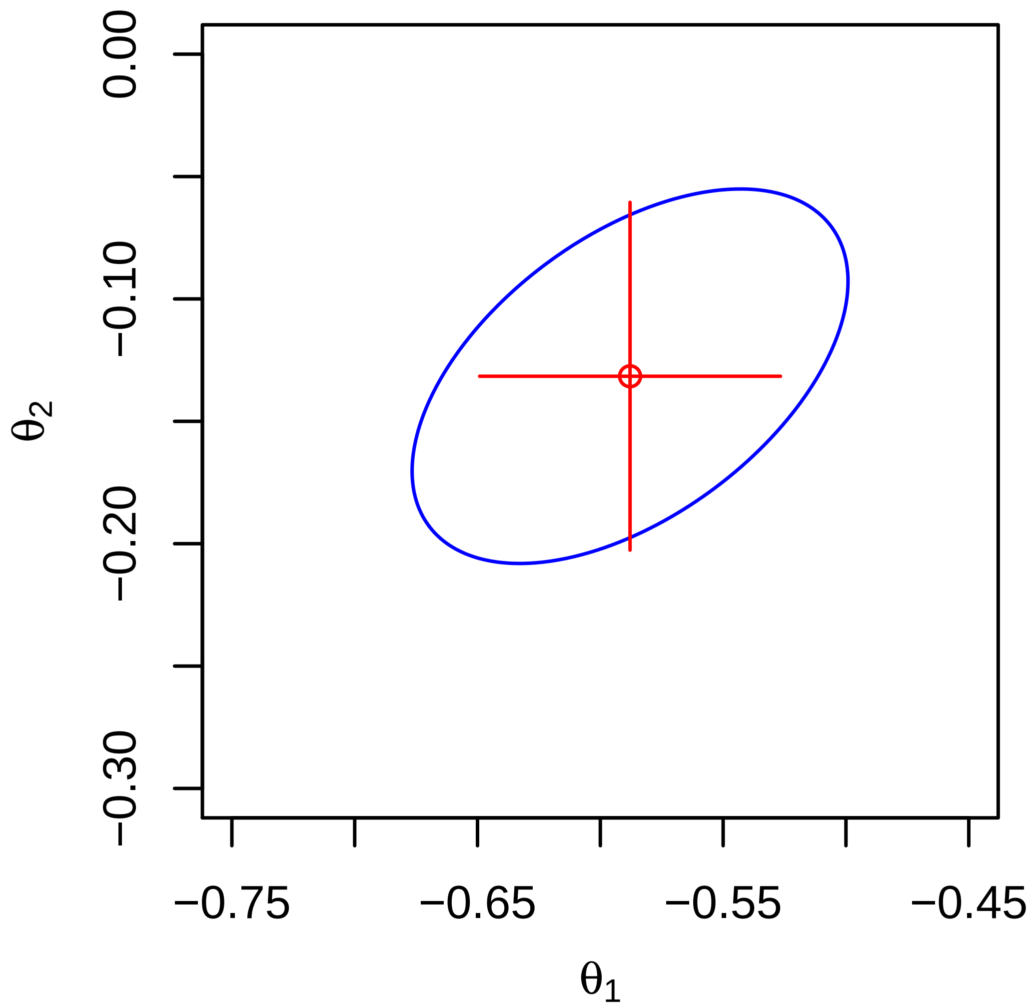
Innovations Algorithm for Atomic Clock

- appears to have converged by $n = 15$ (maybe sooner?)
- base estimates & 95% CIs for $\theta_1, \ldots, \theta_4$ on $\hat{\theta}_{15,1}, \ldots, \hat{\theta}_{15,4}$:

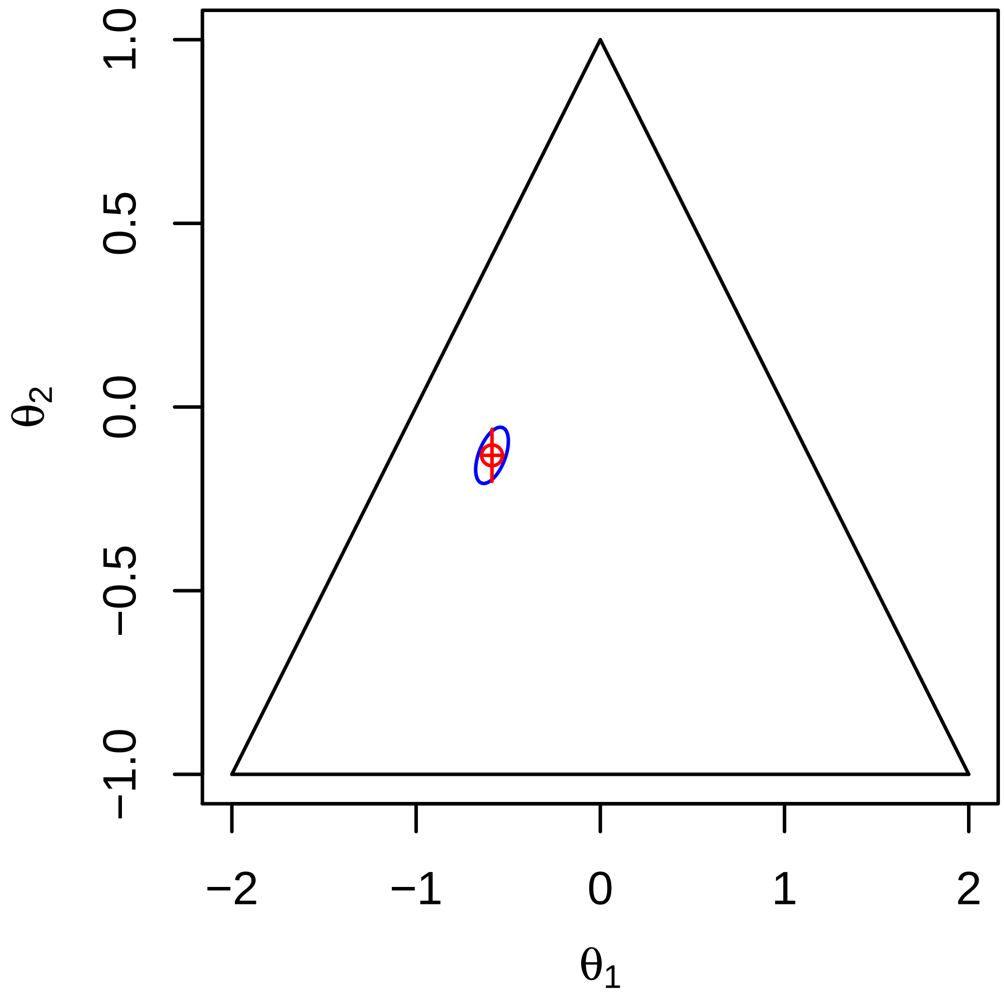
j	$\hat{\theta}_j$	lower bound	upper bound
1	-0.5879	-0.6491	-0.5266
2	-0.1316	-0.2026	-0.0605
3	-0.0690	-0.1405	0.0025
4	-0.0044	-0.0760	0.0672

- moment matching for MA(1) model gave $\hat{\theta} \doteq -0.4352$, which is *not* within 95% CI based on $\hat{\theta}_1$
- CIs suggest MA(2) model is appropriate
- here $\hat{\sigma}^2 = \hat{v}_{15} \doteq 20.782$ (got $\hat{\sigma}^2 \doteq 23.919$ from MA(1) moment matching)

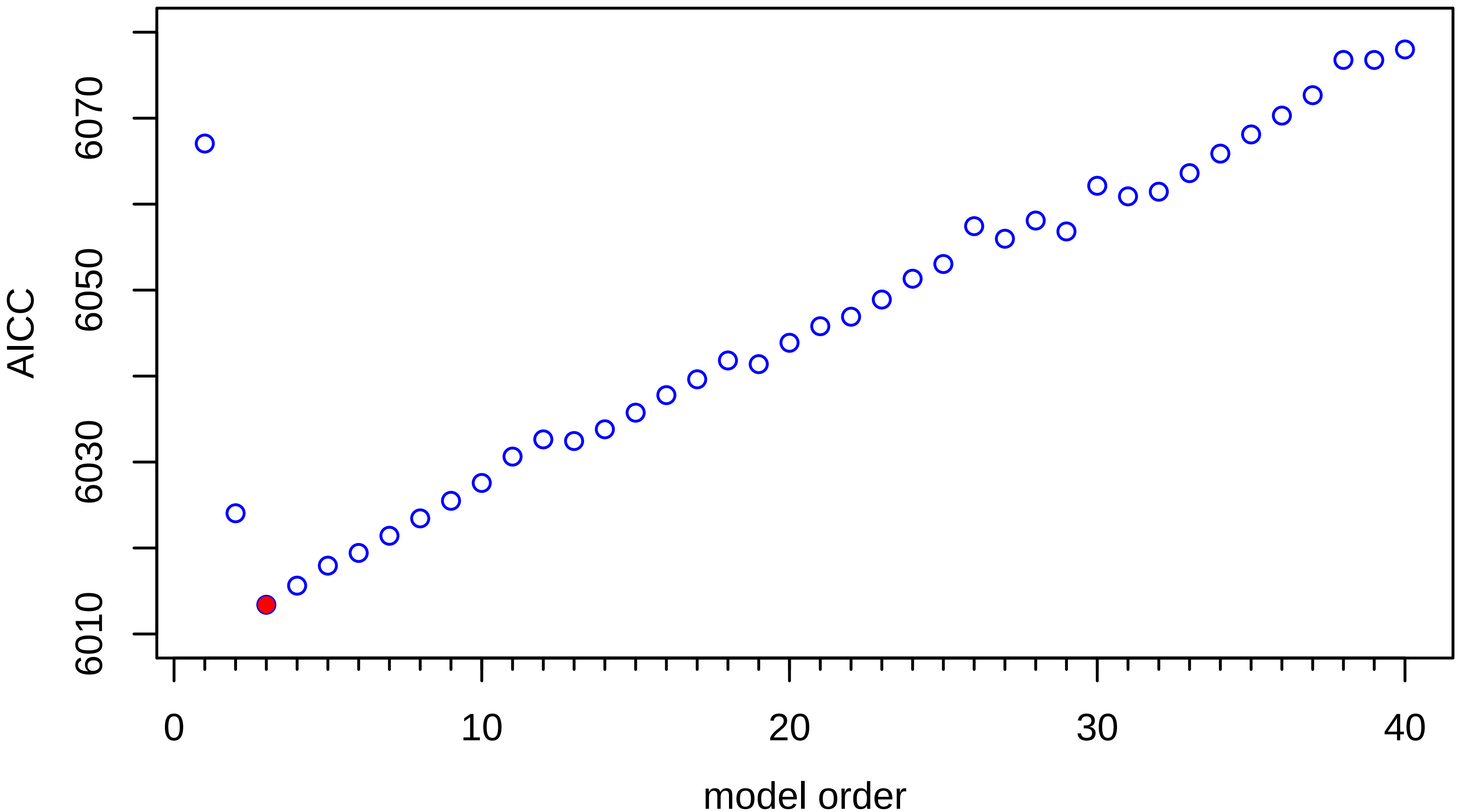
95% Confidence Region for $\boldsymbol{\theta}$



95% Confidence Region & Invertibility Region for $\boldsymbol{\theta}$



AICC for Atomic Clock Series



Parameter Estimation for Mixed ARMA(p, q) Models

- so far have discussed estimation techniques appropriate for pure AR(p) models and pure MA(q) models
- will refer to ARMA model for which both $p > 0$ and $q > 0$ as a 'mixed' ARMA model
  - note: ARMA processes are presumed to be causal (& hence stationary) and invertible
- can handle mixed ARMA models using
  - innovations algorithm
  - higher-order Yule–Walker method with innovations algorithm
  - Hannan–Rissanen algorithm, an example of a so-called least squares estimator

Innovations Algorithm for Mixed ARMA Models: I

- assume causal and invertible ARMA process:

  $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$,

  where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- because of causality, can write

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

where $\psi_0 = 1$ and

$$\psi_j = \theta_j + \sum_{i=1}^{\min\{p,j\}} \phi_i \psi_{j-i}, \quad j \geq 1, \qquad (*)$$

for which we define $\theta_j = 0$ for $j > q$ (see overhead VIII–16)

Innovations Algorithm for Mixed ARMA Models: II

- knowing $\psi_1 \ldots, \psi_{p+q}$, can solve for ϕ_i's & θ_j's, as follows
- for $j = 1, \ldots, q$, equation $(*)$ involves $\theta_1, \ldots, \theta_q$ directly:

$$\psi_j = \theta_j + \sum_{i=1}^{\min\{p,j\}} \phi_i \psi_{j-i} \qquad (*)$$

- for $j = q+1, \ldots, q+p$, $(*)$ does not involve θ_j's directly:

$$\psi_j = \sum_{i=1}^{\min\{p,j\}} \phi_i \psi_{j-i} \qquad (\dagger)$$

- use $(\dagger)$ to solve for ϕ_i's, after which $(*)$ gives θ_j's:

$$\theta_j = \psi_j - \sum_{i=1}^{\min\{p,j\}} \phi_i \psi_{j-i}$$

Innovations Algorithm for Mixed ARMA Models: III

- IA takes ACVF and gives $\theta_{m,j}$'s and v_m's, where $\theta_{m,j} \to \psi_j$ and $v_m \to \sigma^2$ as $m \to \infty$
- to get estimates of ϕ_i's and θ_j's,
  1. use IA with sample ACVF to get estimates of $\theta_{m,j}$ (with m chosen large enough to ensure convergence – tricky!)
  2. set $\hat{\psi}_j$ equal to estimate of $\theta_{m,j}$
  3. use $\hat{\psi}_j$'s with p equations to get $\hat{\phi}_i$'s
  4. use $\hat{\psi}_j$'s and $\hat{\phi}_i$'s with q additional equations to get $\hat{\theta}_j$'s
- use of IA with sample ACVF also yields $\hat{v}_m$ as an estimator of σ^2 (again need to select m large enough to ensure convergence).

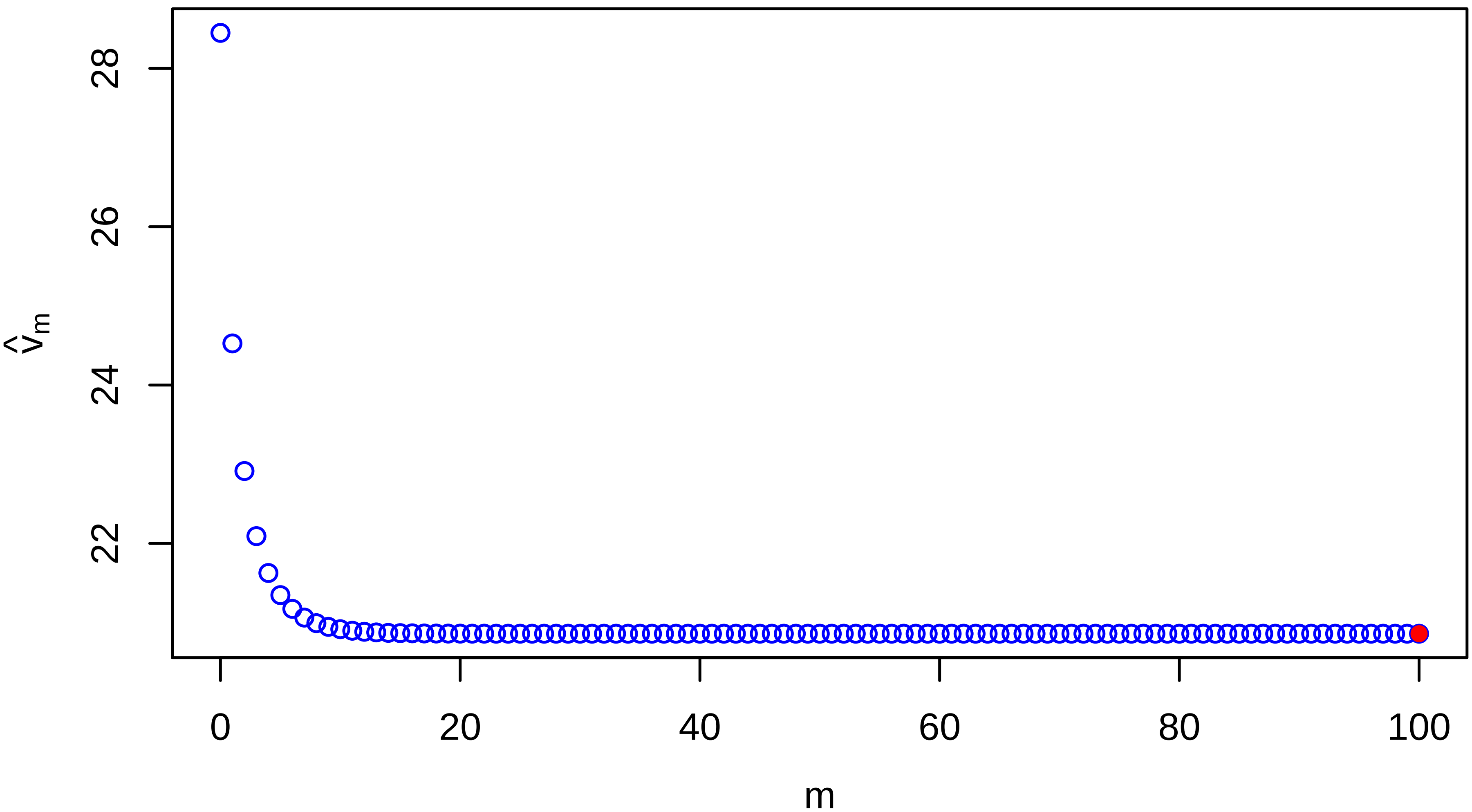
Innovations Algorithm for Mixed ARMA Models: IV

- B&D note that
  - resulting $\hat{\phi}_i$ need *not* correspond to a causal process
  - order selection using sample ACVF and PACF dicey: no clear patterns distinguishing, e.g., ARMA(2,1) & ARMA(1,2)
  - order selection can still be done using AICC

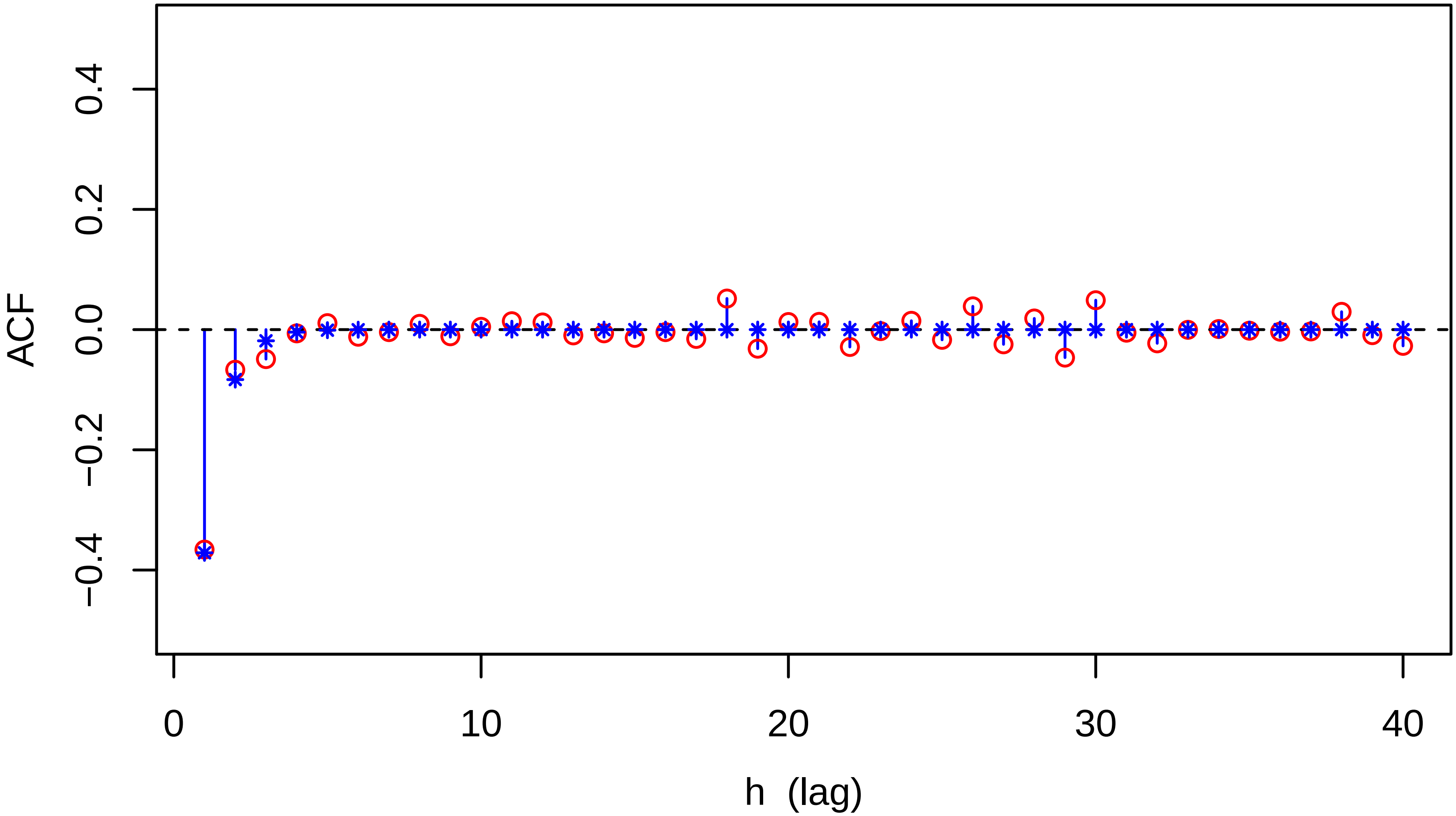
Example – Atomic Clock Series: I

- as an example, let's use IA to fit an ARMA(1,1) model
$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$
to atomic clock series
- the $p + q = 2$ relevant equations are
$$\psi_1 = \theta + \phi \;\&\; \psi_2 = \phi\psi_1, \text{ yielding } \phi = \frac{\psi_2}{\psi_1} \;\&\; \theta = \psi_1 - \phi$$
- basing our estimates of ψ_1 and ψ_2 on
$$\hat{\theta}_{15,1} \doteq -0.5879 \text{ and } \hat{\theta}_{15,2} \doteq -0.1316$$
(called $\hat{\theta}_1$ and $\hat{\theta}_2$ on overhead XIII–76) yields
$$\hat{\phi} \doteq 0.2238 \text{ and } \hat{\theta} \doteq -0.8117$$
(corresponds to a causal and invertible ARMA(1,1) model)
- get $\hat{\sigma}^2 = 20.860$ (compared to $\hat{\sigma}^2 \doteq 20.782$ for MA(2) model)

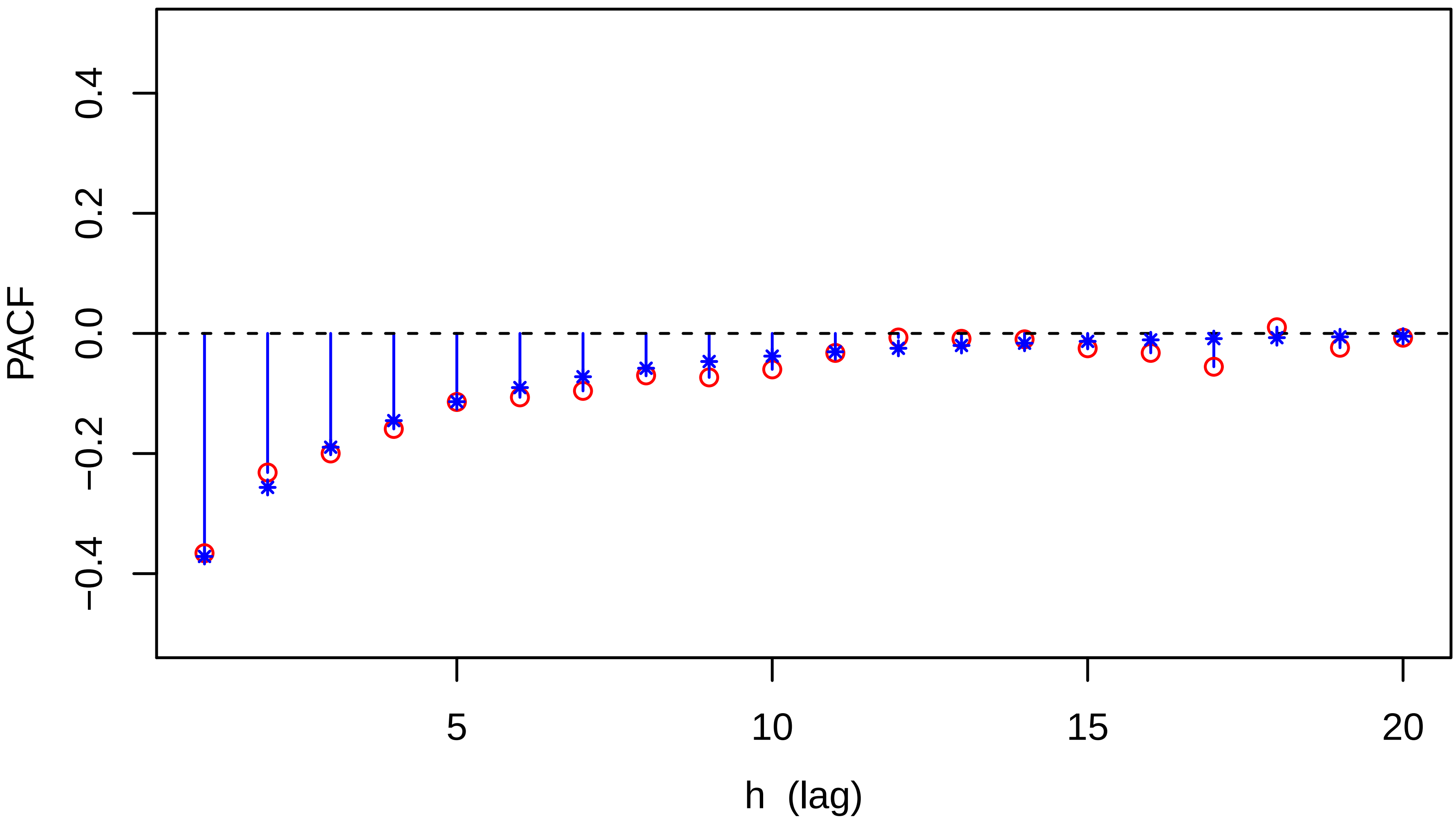
Estimation of σ^2 via $\hat{v}_{100}$ from Innovations Algorithm



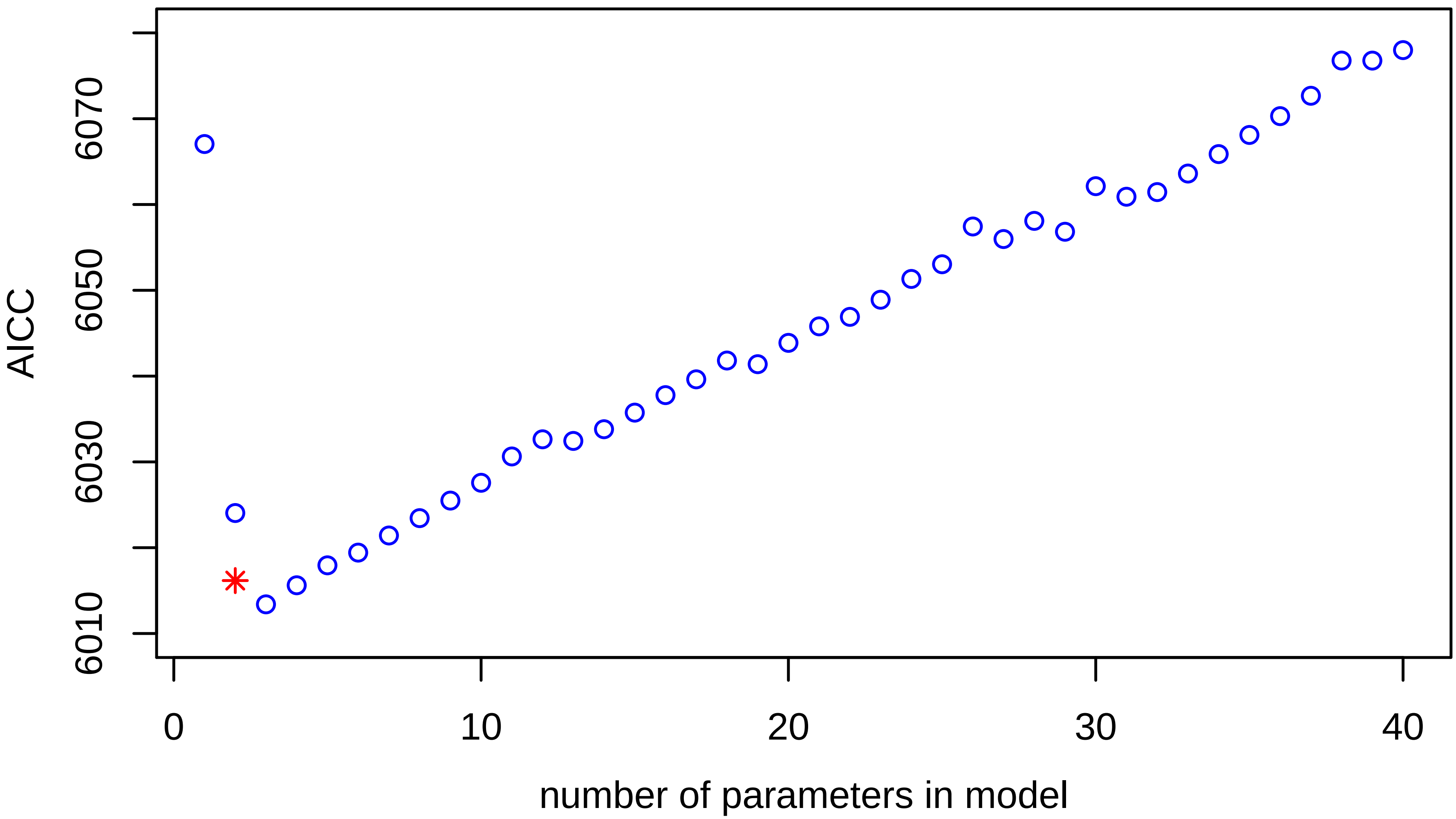
Sample & ARMA(1,1) ACF for Atomic Clock



Sample & ARMA(1,1) PACF for Atomic Clock



AICC for Atomic Clock Series



Higher-Order Yule–Walker Method: I

- alternative to IA algorithm for handling mixed ARMA models is based on structure of ACVF $\{\gamma(h)\}$ for such models
- as noted before (overhead IX–20), ARMA(p, q) ACVF satisfies

$$\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p) = 0$$

  for all $k \geq q + 1$
- note: above equation does *not* involve MA coefficients
- can use so-called higher-order Y–W equations to get ϕ_i's without interference from MA coefficients:

$$\begin{aligned}
\phi_1\gamma(q) + \phi_2\gamma(q-1) + \cdots + \phi_p\gamma(q-p+1) &= \gamma(q+1) \\
\phi_1\gamma(q+1) + \phi_2\gamma(q) + \cdots + \phi_p\gamma(q-p+2) &= \gamma(q+2) \\
&\vdots \\
\phi_1\gamma(q+p-1) + \phi_2\gamma(q+p-2) + \cdots + \phi_p\gamma(q) &= \gamma(q+p)
\end{aligned}$$

Higher-Order Yule–Walker Method: II

- with ϕ_i's known, can filter time series $X_1, \ldots, X_n$ and get output $Y_{p+1}, \ldots, Y_n$ with MA(q) structure:

$$\begin{aligned} Y_t &\stackrel{\text{def}}{=} X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} \\ &= Z_t + \theta_1 Z_{q-1} + \cdots + \theta_q Z_{t-q} \end{aligned}$$

- higher-order Y–W method with IA thus consists of
  - substituting $\hat{\gamma}(h)$'s into higher-order Y–W equations and solving to get estimates $\hat{\phi}_i$
  - using $\hat{\phi}_i$'s to filter time series to get output, say Y_t'
  - forming sample ACVF for Y_t''s and using these as input to IA to estimate MA coefficients θ_j

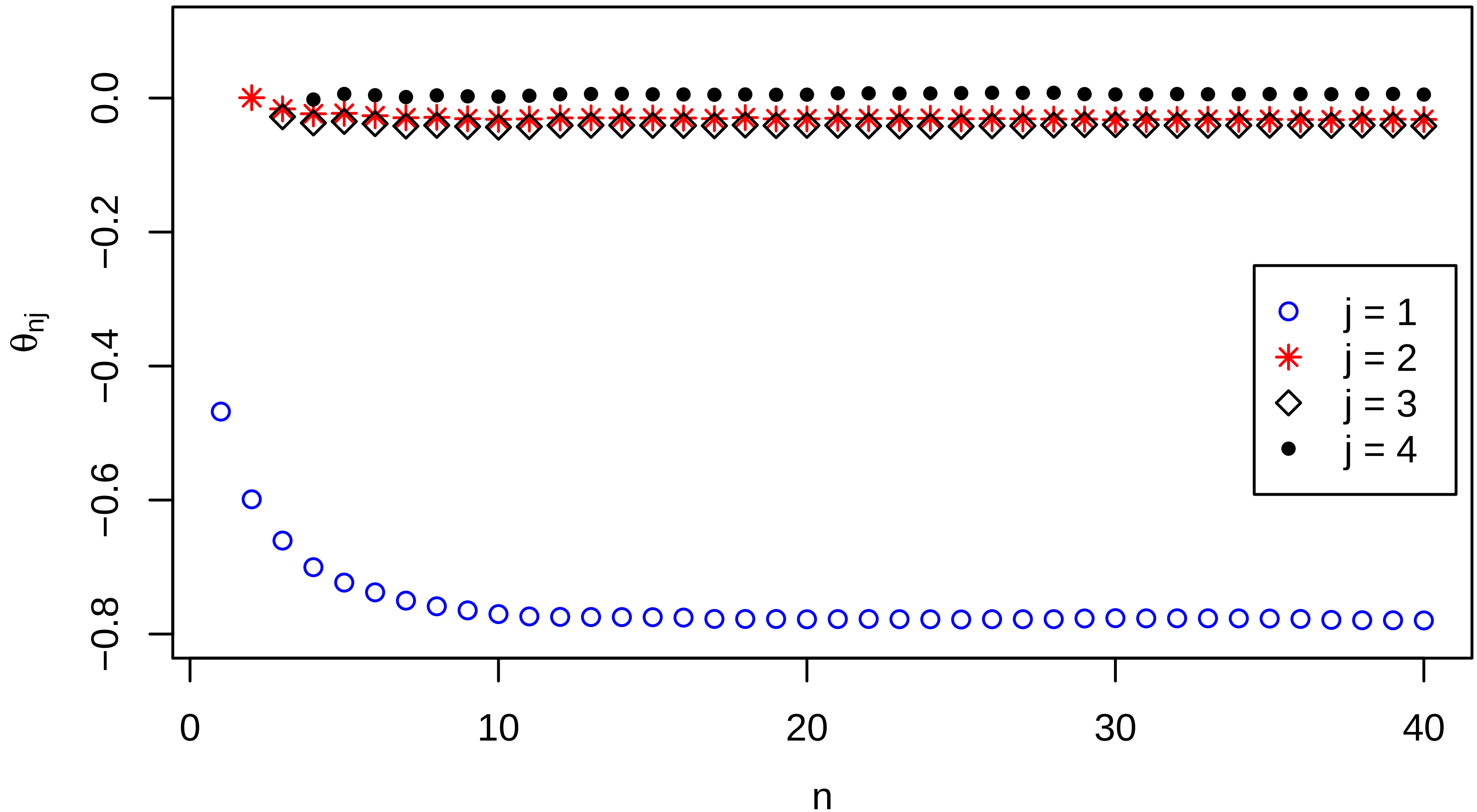
Example – Atomic Clock Series: II

- as an example, let's use scheme to fit an ARMA(1,1) model to atomic clock series
- relevant higher-order Y–W equation is $\phi\gamma(1) = \gamma(2)$, yielding

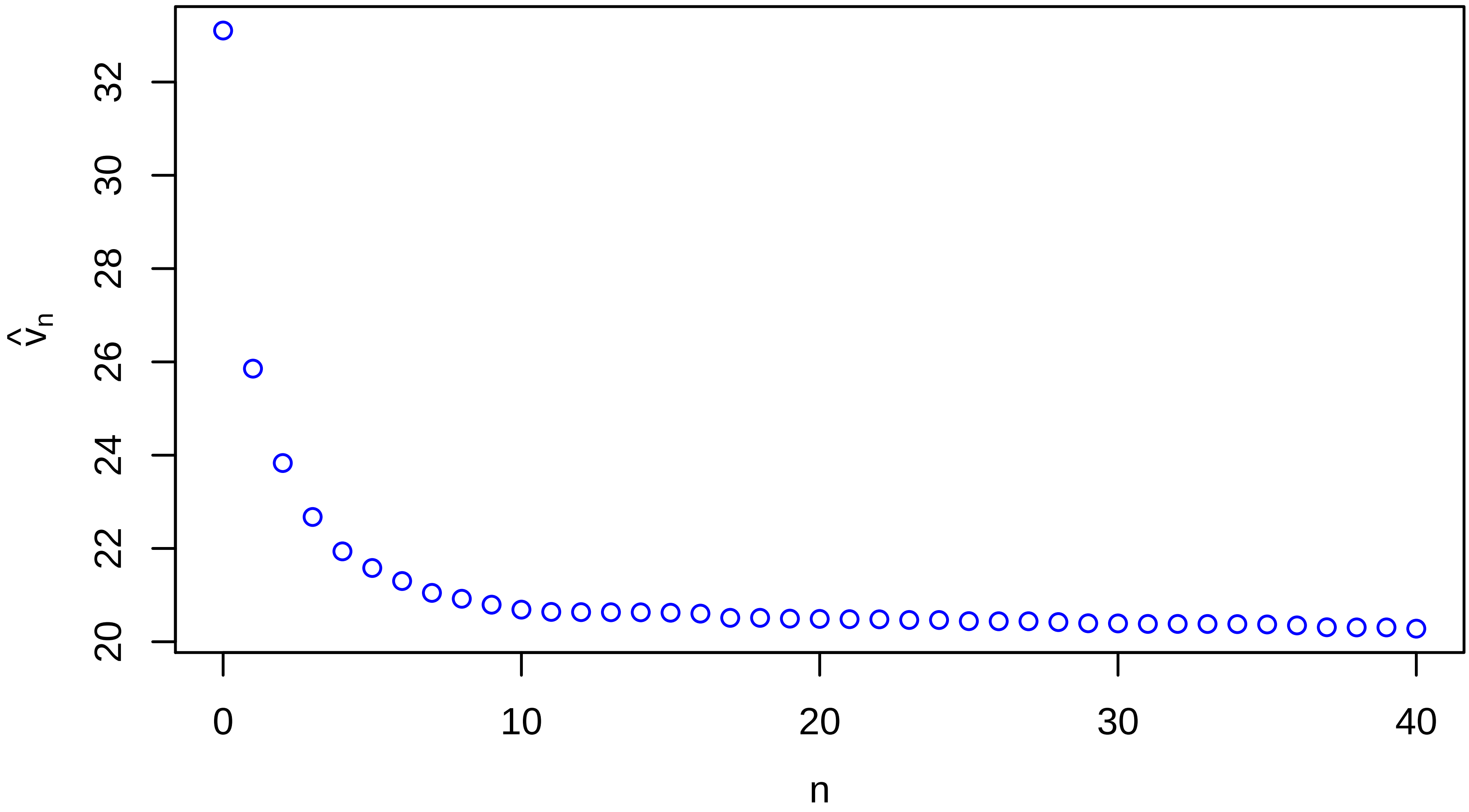
$$\hat{\phi} = \frac{\hat{\gamma}(2)}{\hat{\gamma}(1)} \doteq 0.1823, \text{ as compared to } \hat{\phi} \doteq 0.2238 \text{ using IA}$$

- estimate corresponds to a causal process, but might not happen for other time series (no reason why $\hat{\gamma}(1) \approx 0$ can't occur)
- forming sample ACVF for $Y_t' = X_t - \hat{\phi}X_{t-1}$, $t = 2, \ldots, n$, and feeding it into IA yields $\hat{\theta}_{n,j}$'s and $\hat{v}_n$'s shown on next overheads

Convergence of $\hat{\theta}_{n,j}$'s for Y_t''s

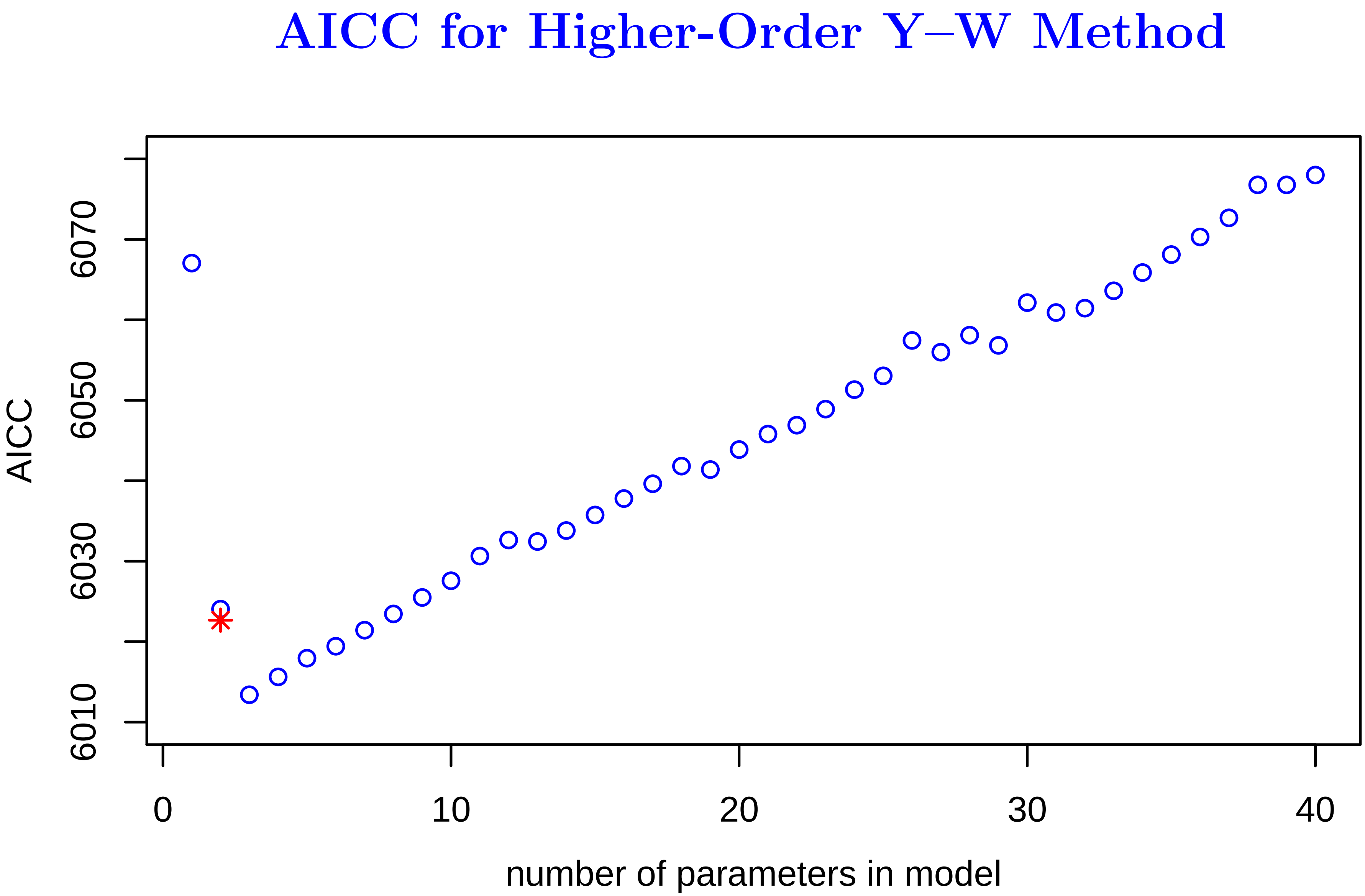


Convergence of $\hat{v}_n$'s for Y_t''s



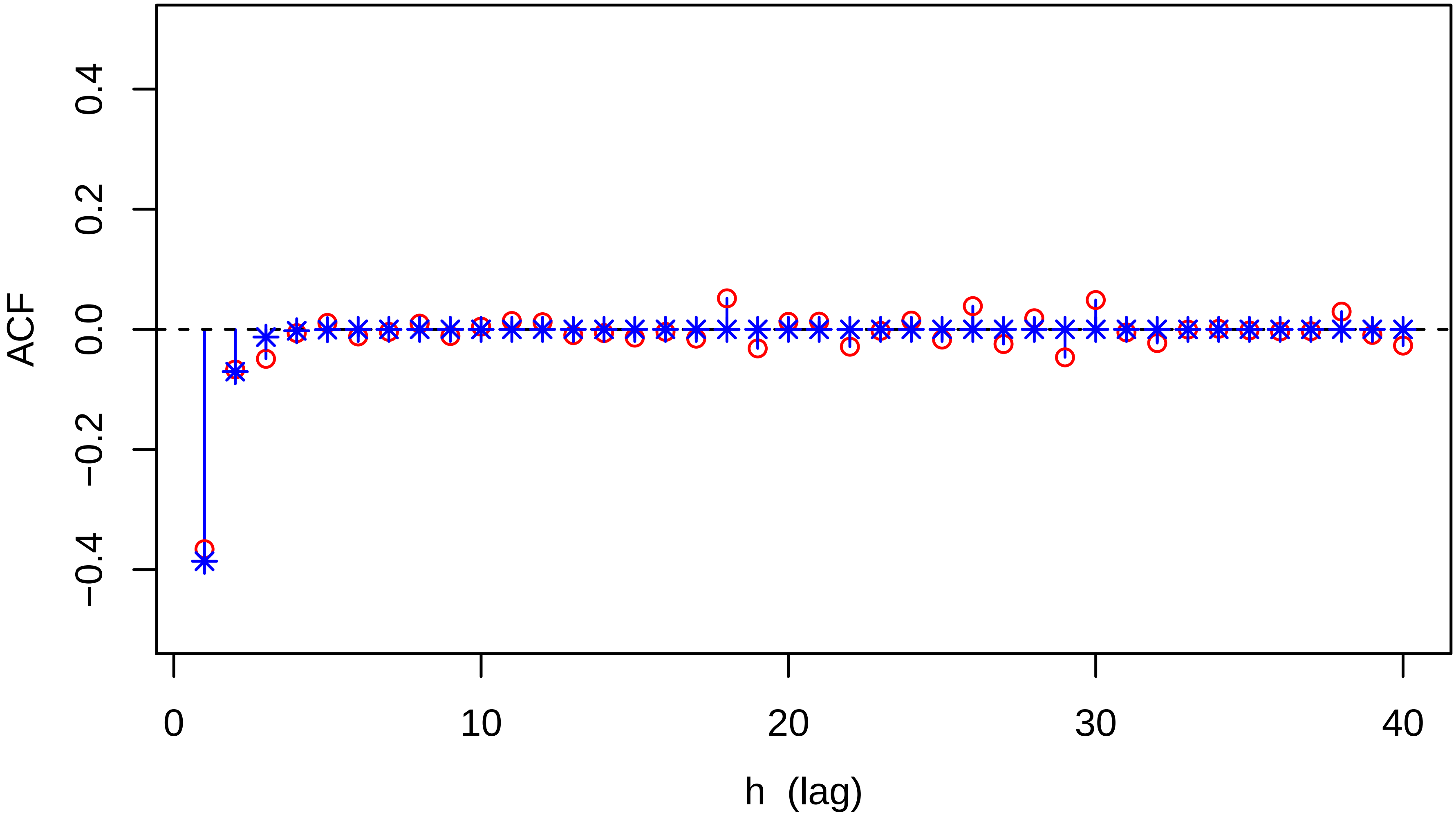
Example – Atomic Clock Series: III

- using $\hat{\theta}_{15,1}$ to estimate θ in ARMA(1,1) model, get $\hat{\theta} \doteq -0.7748$ compared to $\hat{\theta} \doteq -0.8117$ using IA by itself (overhead XIII–85)
- using $\hat{v}_{15}$ to estimate σ^2 yields $\hat{\sigma}^2 \doteq 20.631$ as compared to IA-based $\hat{\sigma}^2 \doteq 20.860$
- sampling theory for $\hat{\theta}_{n,j}$'s suggests that those for $j = 2, 3$ and 4 are not significantly different from zero; i.e., ARMA(1,q) model with $q > 1$ not indicated
- AICC for fitted ARMA(1,1) model is 6022.7, so model is less likely than IA-based model with AICC of 6016.2
- next overheads
  - show AICC compared to ones for IA-based MA(q) models
  - compare theoretical and sample ACVFs and PACFs

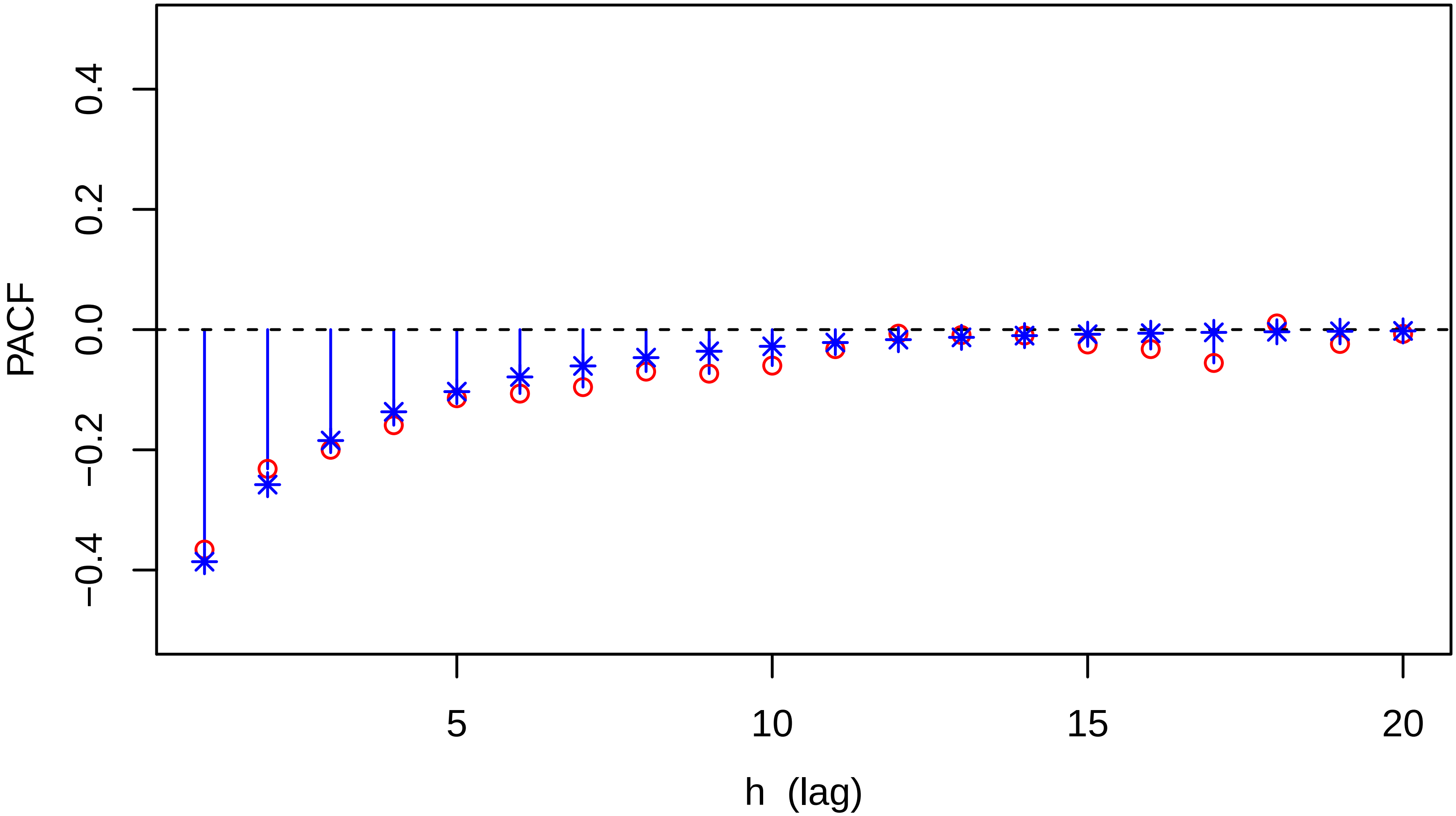


AICC for Higher-Order Y–W Method
AICC
6010
6030
6050
6070
0
10
20
30
40
number of parameters in model

Sample & ARMA(1,1) ACF for Atomic Clock



Sample & ARMA(1,1) PACF for Atomic Clock



Least Squares Estimators: I

- as prelude to Hannan–Rissanen algorithm, consider least squares (LS) estimator for AR(p) coefficients
- express AR(p) model as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

- above looks like a multiple regression model, except explanatory variables (predictors) are lagged versions of dependent variable (response) X_t

Least Squares Estimators: II

- given time series $X_1, \ldots, X_n$, have $n-p$ observations for which we can also get required explanatory variables:

$$
\begin{aligned}
X_{p+1} &= \phi_1 X_p \quad + \cdots + \phi_p X_1 \quad + Z_{p+1} \\
X_{p+2} &= \phi_1 X_{p+1} + \cdots + \phi_p X_2 \quad + Z_{p+2} \\
&\vdots \\
X_n &= \phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} + Z_n
\end{aligned}
$$

- in matrix formulation, can write as $\boldsymbol{X}_{n:p+1} = A\boldsymbol{\phi} + \boldsymbol{Z}$, where

$$
\boldsymbol{X}_{n:p+1} = \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_{p+1} \end{bmatrix}, A = \begin{bmatrix} X_{n-1} & \cdots & X_{n-p} \\ X_{n-2} & \cdots & X_{n-p-1} \\ \vdots & \vdots & \vdots \\ X_p & \cdots & X_1 \end{bmatrix}, \boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} Z_n \\ Z_{n-1} \\ \vdots \\ Z_{p+1} \end{bmatrix}
$$

Least Squares Estimators: III

- by definition, ordinary LS estimator of $\boldsymbol{\phi}$ minimizes

$$S_f(\boldsymbol{\phi}) \stackrel{\text{def}}{=} \sum_{t=p+1}^{n} \left(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}\right)^2$$

(note: rationale for subscript 'f' forthcoming!)

- S_f is a quadratic function of $\boldsymbol{\phi}$, so any minimizing $\boldsymbol{\phi}$ must satisfy, for $i = 1, \ldots, p$,

$$\frac{\partial S_f(\boldsymbol{\phi})}{\partial \phi_i} = -2 \sum_{t=p+1}^{n} \left(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}\right) X_{t-i} = 0$$

- yields set of p so-called normal equations:

$$\sum_{t=p+1}^{n} \phi_1 X_{t-1} X_{t-i} + \cdots + \phi_p X_{t-p} X_{t-i} = \sum_{t=p+1}^{n} X_t X_{t-i}$$

Least Squares Estimators: IV

- in matrix formulation, normal equations become $A'A\boldsymbol{\phi} = A'\boldsymbol{X}_{n:p+1}$
- denote solution as $\hat{\boldsymbol{\phi}}_f$ – this is the LS estimator of $\boldsymbol{\phi}$
- $\hat{\boldsymbol{\phi}}_f$ need *not* correspond to causal AR(p) model, and $A'A$ need *not* be positive definite
- interesting connection between $\hat{\boldsymbol{\phi}}_f$ and Y–W estimator $\hat{\boldsymbol{\phi}}$:
  - take time series and add p zeros before X_1 and p zeros after X_n to create a time series, say $\widetilde{X}_1, \ldots, \widetilde{X}_{n+2p}$
  - let $\widetilde{A}$ be analog of A for zero-padded series; then $\widetilde{A}'\widetilde{A}\boldsymbol{\phi} = \widetilde{A}'\widetilde{\boldsymbol{X}}_{n+2p:p+1}$ reduces to $\hat{\Gamma}_p\boldsymbol{\phi} = \hat{\boldsymbol{\gamma}}_p$; i.e., LS estimator based on $\{\widetilde{X}_t\}$ is *identical* to Y–W estimator based on $\{X_t\}$!
  - $\hat{\boldsymbol{\phi}}$ corresponds to causal AR(p) model, and $\widetilde{A}'\widetilde{A} = \hat{\Gamma}_p$ is positive definite – adding zeros acts as regularization procedure!

Least Squares Estimators: V

- in view of

$$S_f(\boldsymbol{\phi}) = \sum_{t=p+1}^{n} \left(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}\right)^2,$$

  can regard $\hat{\boldsymbol{\phi}}_f$ as arising from minimization of sum of squared forward prediction errors (rationale for subscript 'f')

- in the same spirit as Burg's algorithm, can also consider backward prediction errors:

$$S_b(\boldsymbol{\phi}) = \sum_{t=1}^{n-p} \left(X_t - \phi_1 X_{t+1} - \cdots - \phi_p X_{t+p}\right)^2$$

- leads to forward/backward LS estimator $\hat{\boldsymbol{\phi}}_{fb}$, which is the value of $\boldsymbol{\phi}$ minimizing

$$S_f(\boldsymbol{\phi}) + S_b(\boldsymbol{\phi})$$

Hannan–Rissanen Algorithm: I

- Hannan–Rissanen (H–R) algorithm extends LS estimator to work with ARMA(p, q) processes
- reexpress ARMA model to mimic a multiple regression model:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \\ &\qquad + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2) \end{aligned}$$

- explanatory variables now also include lagged versions of errors Z_t — alas, these can't be directly observed
- H–R gets around this problem by creating surrogates $\hat{Z}_t$ for unobservable Z_t's and then forging ahead with LS procedure

Hannan–Rissanen Algorithm: II

- start by fitting high-order AR(m) model to time series $X_1, \ldots, X_n$ using Y–W, where $m \gg \max\{p, q\}$ (but also need $m \ll n$)
- hope is that high-order AR model can closely mimic covariance structure of parsimonious low-order ARMA model
- estimate Z_t using

  $$\hat{Z}_t = X_t - \hat{\phi}_{m,1} X_{t-1} - \cdots - \hat{\phi}_{m,m} X_{t-m}, \quad t = m+1, \ldots, n$$

- estimate $\boldsymbol{\beta} = [\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]'$ by minimizing

$$S(\boldsymbol{\beta}) = \sum_{t=m+1+q}^{n} \Big(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta_1 \hat{Z}_{t-1} - \cdots - \theta_q \hat{Z}_{t-q} \Big)^2$$

- let $\hat{\boldsymbol{\beta}} = [\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q]'$ denote resulting estimator, which H & R advocate polishing up as follows

Hannan–Rissanen Algorithm: III

- to polish estimates of white noise component, recursively set

$$\tilde{Z}_t = \begin{cases} 0, & 1 \le t \le \max\{p, q\}; \\ X_t - \sum_{i=1}^{p} \hat{\phi}_i X_{t-i} - \sum_{j=1}^{q} \hat{\theta}_j \tilde{Z}_{t-j}, & \max\{p, q\} < t \le n \end{cases}$$

- since $\tilde{Z}_t$'s might not be quite white, recursively set

$$V_t = \begin{cases} 0, & 1 \le t \le \max\{p, q\}; \\ \sum_{i=1}^{p} \hat{\phi}_i V_{t-i} + \tilde{Z}_t, & \max\{p, q\} < t \le n \end{cases}$$

and

$$W_t = \begin{cases} 0, & 1 \le t \le \max\{p, q\}; \\ -\sum_{j=1}^{q} \hat{\theta}_j W_{t-j} + \tilde{Z}_t, & \max\{p, q\} < t \le n \end{cases}$$

- note: $\hat{\phi}(B)V_t = \tilde{Z}_t$ & $\hat{\theta}(B)W_t = \tilde{Z}_t$, so $\hat{\phi}(B)V_t = \hat{\theta}(B)W_t$

Hannan–Rissanen Algorithm: IV

- estimate $\boldsymbol{\beta} = [\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]'$ by minimizing

$$S(\boldsymbol{\beta}) = \sum_{t=\max\{p,q\}+1}^{n} \left(\tilde{Z}_t - \sum_{i=1}^{p} \phi_i V_{t-i} - \sum_{j=1}^{q} \theta_j W_{t-j} \right)^2$$

- let $\hat{\boldsymbol{\beta}}^{\dagger} = [\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q]'$ denote resulting estimator
- H–R estimator is $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^{\dagger}$, but B&D stick with just $\hat{\boldsymbol{\beta}}$
- three comments
  1. can handle both pure MA(q) & mixed ARMA(p, q) models
  2. usual formulation of H–R calls for use of Y–W, but Burg is a better choice (in particular, $\hat{Z}_t$'s are computed as part of Burg's algorithm)
  3. as in IA, choice of $m > \max\{p, q\}$ requires some care

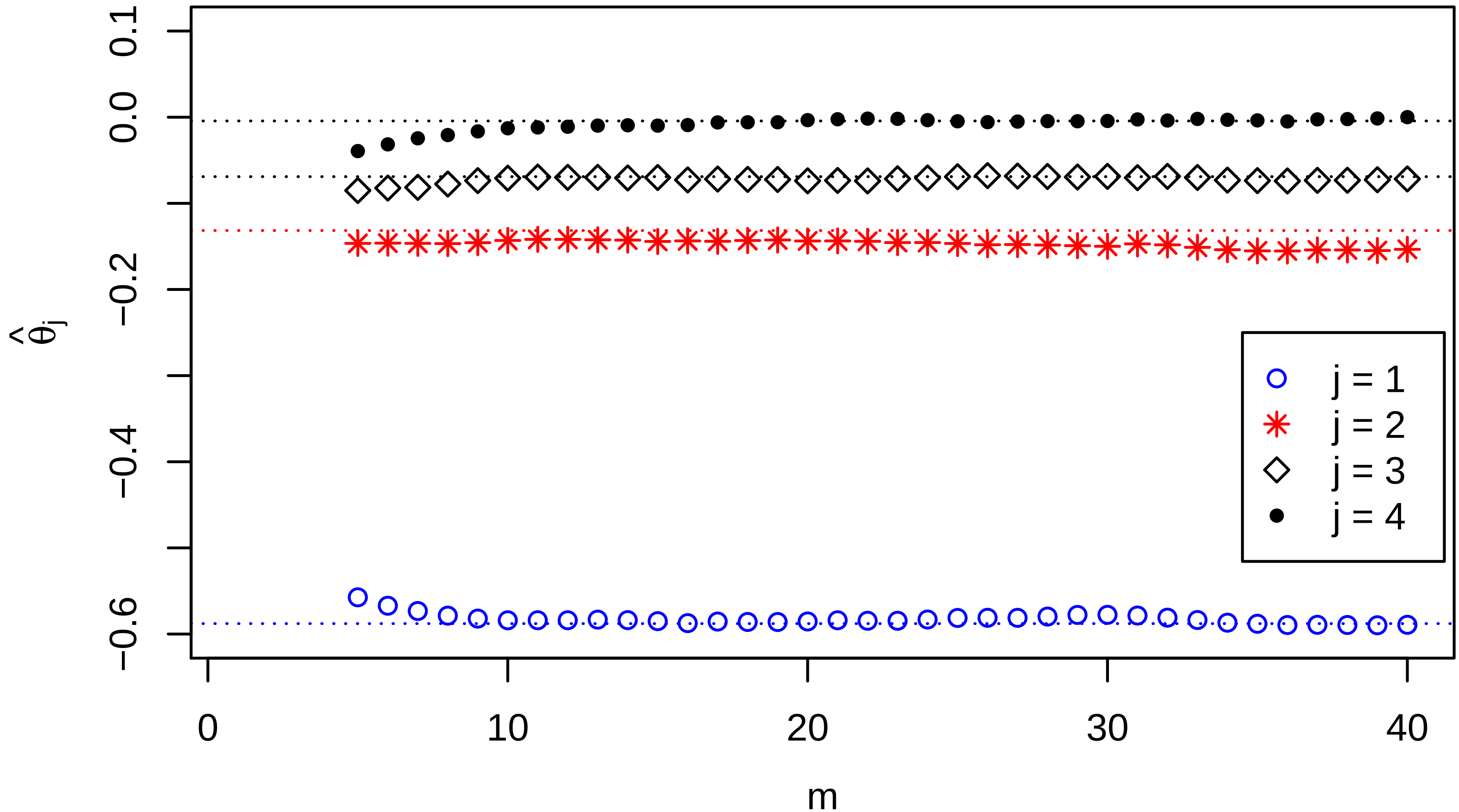
Example – Atomic Clock Series: IV

- as an example, let's use H–R to fit an MA(4) model to atomic clock series to compare with IA results
- next two overheads look at
  1. dependence of estimates $\hat{\theta}_1, \ldots, \hat{\theta}_4$ on order m for approximating AR process; $m = 5$ to 40; $m = 15$ looks like good choice; dotted lines indicate $\hat{\theta}_j = \hat{\theta}_{15,j}$ from IA
  2. same, but now for refinement $\tilde{\theta}_j$; will use $m = 15$ again

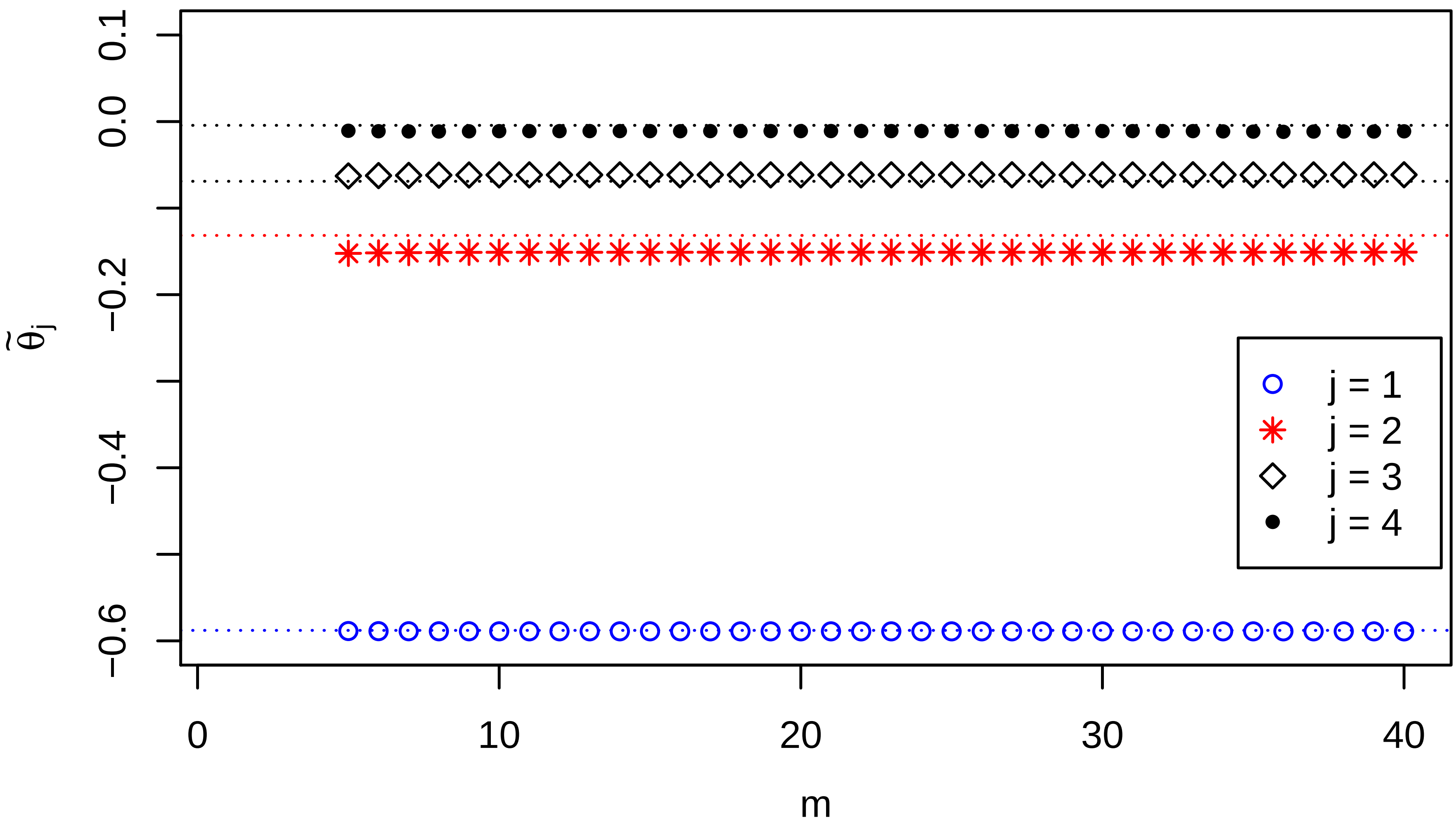
j	$\hat{\theta}_{HR,j}$	$\tilde{\theta}_{HR,j}$	$\hat{\theta}_{IA,j}$	lower bound	upper bound
1	−0.5860	−0.5890	−0.5879	−0.6491	−0.5266
2	−0.1426	−0.1509	−0.1316	−0.2026	−0.0605
3	−0.0723	−0.0616	−0.0690	−0.1405	0.0025
4	−0.0058	−0.0110	−0.0044	−0.0760	0.0672

- note: right-most three columns copied from overhead XIII–76

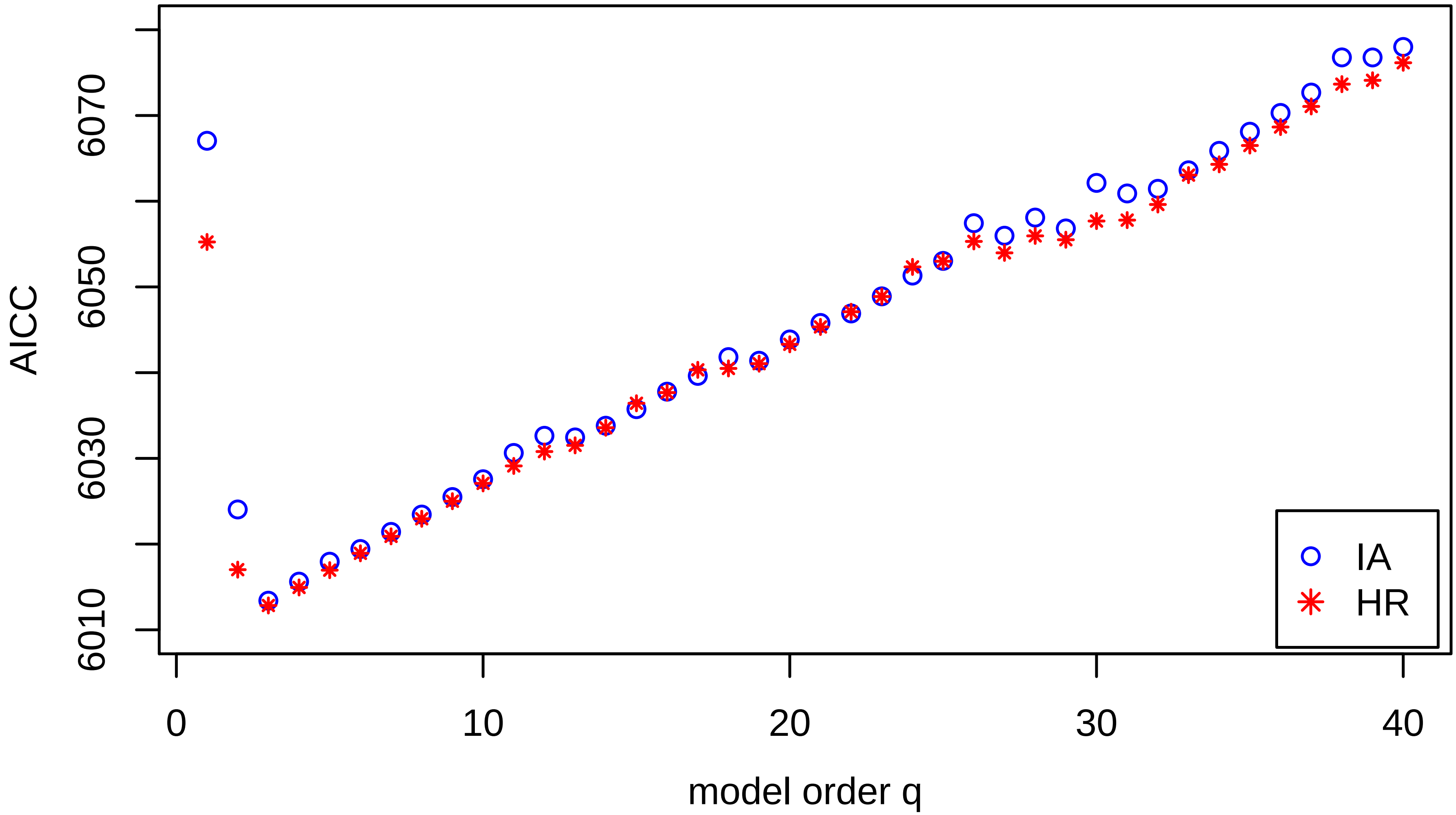
Convergence of $\hat{\theta}_j$'s for Atomic Clock Series



Convergence of $\tilde{\theta}_j$'s for Atomic Clock Series



AICC for MA(q) Models for Atomic Clock Series



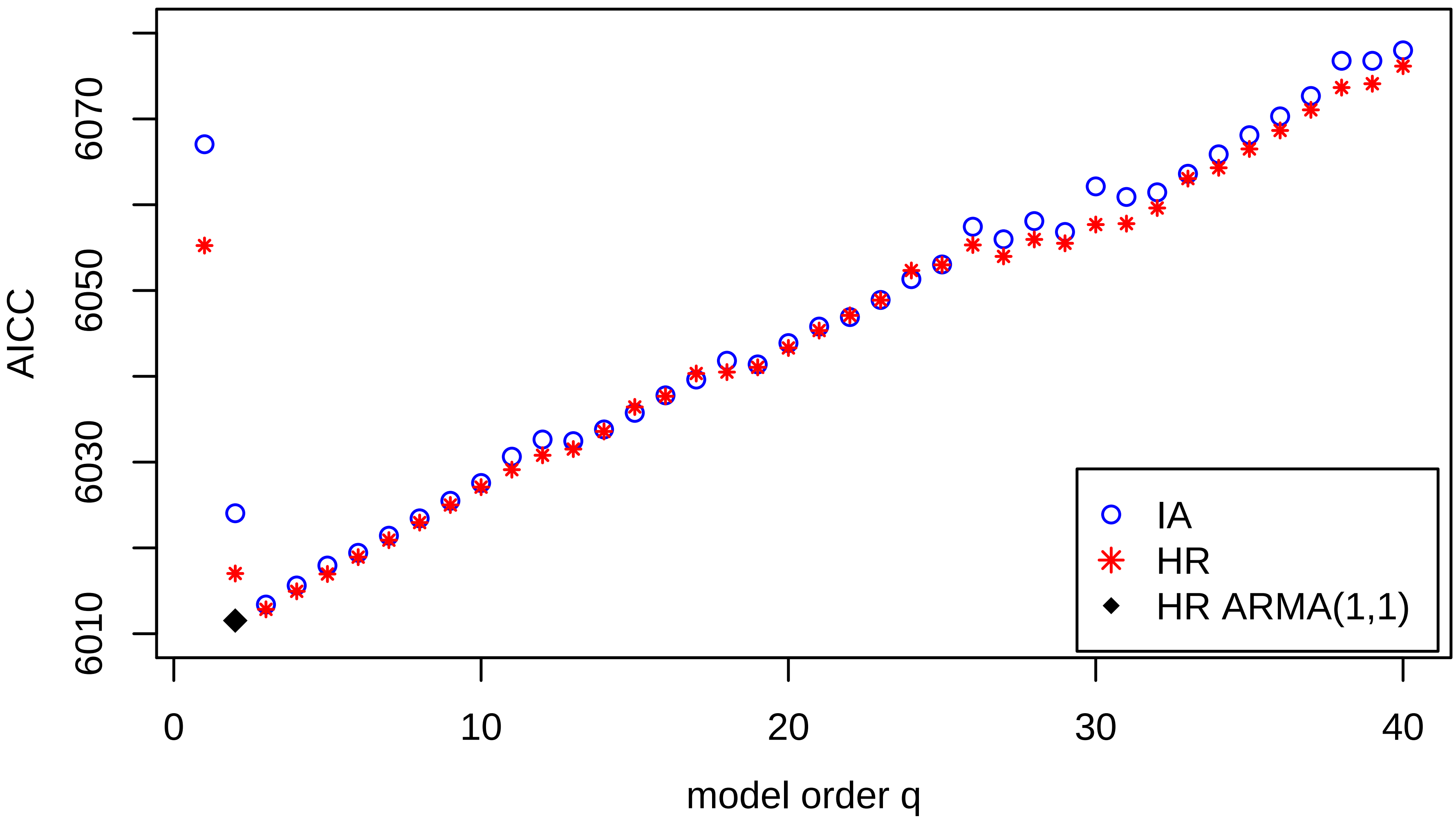
Example – Atomic Clock Series: V

- now let's use H–R to fit an ARMA(1,1) model

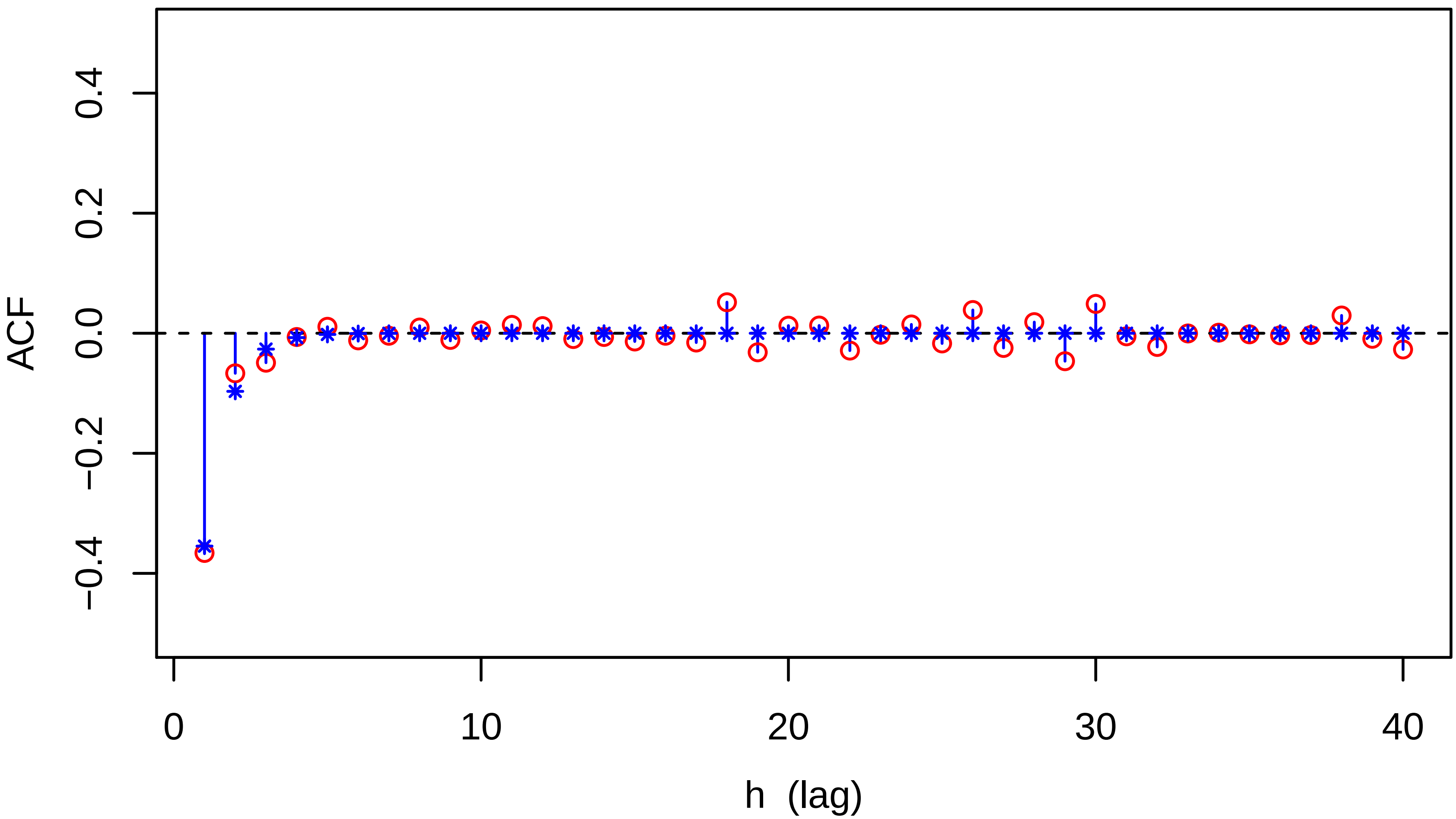
$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

- H–R yields $\hat{\phi} \doteq 0.2730$ and $\hat{\theta} \doteq -0.8662$ $(m = 15)$
- IA yields $\hat{\phi} \doteq 0.2238$ and $\hat{\theta} \doteq -0.8117$
- AICC for IA model is 6016.2, while that for H–R is 6011.5; i.e., H–R model is more likely

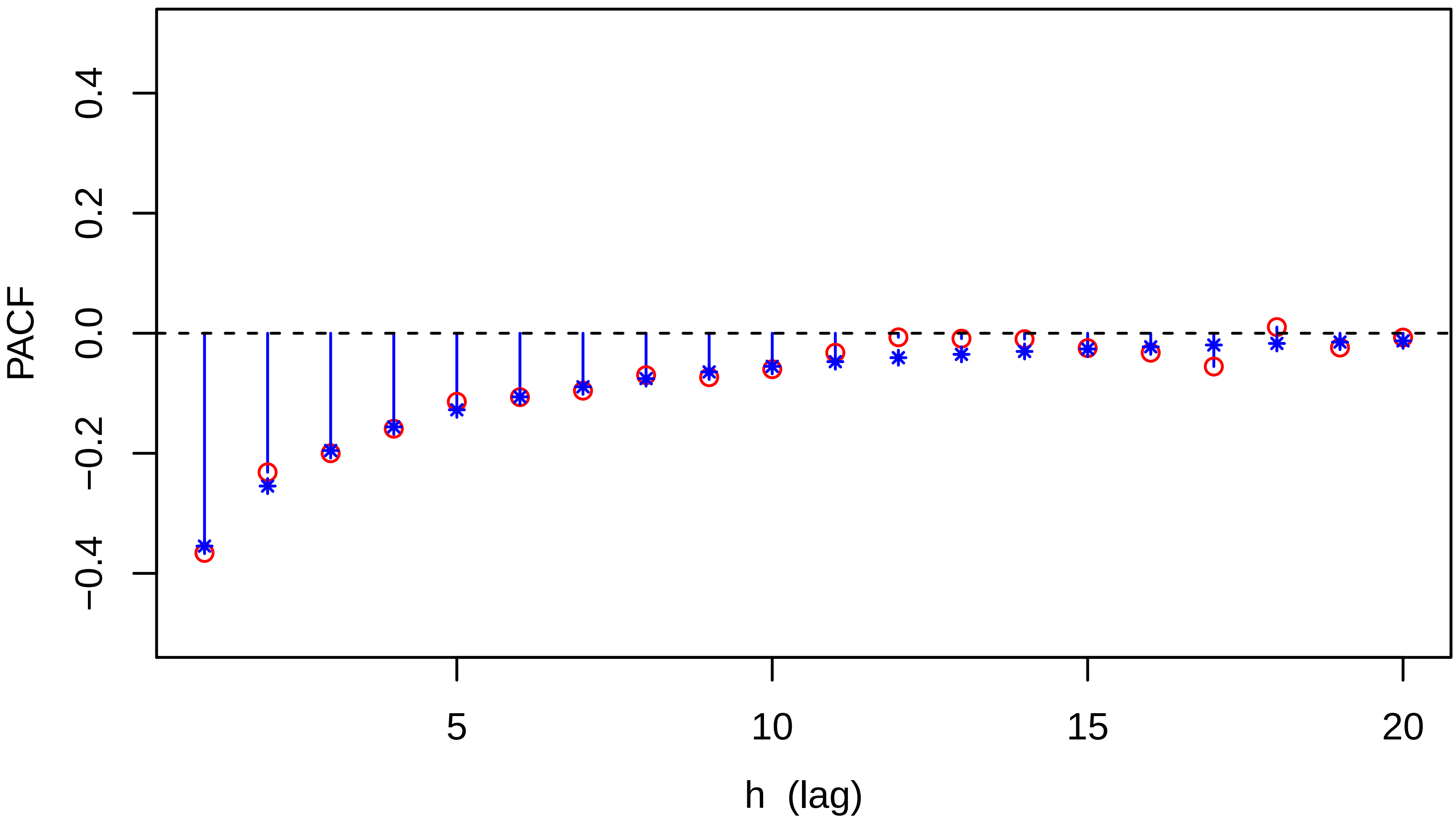
AICC for MA(q) Models for Atomic Clock Series



Sample & ARMA(1,1) ACF for Atomic Clock



Sample & ARMA(1,1) PACF for Atomic Clock



Maximum Likelihood Estimation: I

- as noted before (XIII–22), likelihood for Gaussian zero-mean stationary time series $\boldsymbol{X}_n = [X_n, \ldots, X_1]'$ with covariance matrix Γ_n is given by

$$L(\Gamma_n) = (2\pi)^{-n/2}(\det \Gamma_n)^{-1/2} \exp\left(-\tfrac{1}{2}\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n\right)$$

  and hence

$$-2\ln\left(L(\Gamma_n)\right) = n\ln(2\pi) + \ln(\det \Gamma_n) + \boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n$$

- for ARMA(p, q) time series, parameters $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ & σ^2 set Γ_n
- given $\boldsymbol{X}_n$, can assess likelihood of various parameter settings
- maximum likelihood estimators (MLEs) of parameters are settings such that $L(\Gamma_n)$ is maximized
- note: $L(\Gamma_n)$ is maximized when $-2\ln\left(L(\Gamma_n)\right)$ is minimized

Maximum Likelihood Estimation: II

- key to evaluating $L(\Gamma_n)$ is knowing how to compute $\det \Gamma_n$ and $\boldsymbol{X}_n' \Gamma_n^{-1} \boldsymbol{X}_n$ for various parameter settings
- can appeal to IA equation to get manageable expressions
- key equation behind IA is $\boldsymbol{X}_n = C_n' \boldsymbol{U}_n$, where

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \theta_{n-1,1} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ \theta_{n-1,n-3} & \theta_{n-2,n-4} & \cdots & 1 & 0 & 0 \\ \theta_{n-1,n-2} & \theta_{n-2,n-3} & \cdots & \theta_{2,1} & 1 & 0 \\ \theta_{n-1,n-1} & \theta_{n-2,n-2} & \cdots & \theta_{2,2} & \theta_{1,1} & 1 \end{bmatrix}$$

is an $n \times n$ lower triangular matrix, and $\mathbf{U}_n = [U_n, \ldots, U_1]'$ is a vector of innovations (one-step-ahead prediction errors) – see overheads XI–16 and 17

Maximum Likelihood Estimation: III

- innovations in $\mathbf{U}_n$ are uncorrelated RVs with variances v_{n-1}, $v_{n-2}, \ldots, v_0$
- covariance matrix D_n for $\mathbf{U}_n$ is thus a diagonal matrix, with diagonal elements given by v_j's
- since $\boldsymbol{X}_n = C_n' \boldsymbol{U}_n$, standard result from theory of random vectors (B&D, Equation (A.2.5)) says that covariance Γ_n for $\boldsymbol{X}_n$ can be written as $C_n' D_n C_n$
- can argue (why?) that

$$\det \Gamma_n = (\det C_n')(\det D_n)(\det C_n) = \prod_{j=0}^{n-1} v_j,$$

which gives us a manageable expression for $\det \Gamma_n$

Maximum Likelihood Estimation: IV

- to get a manageable expression for $\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n$, note that, since $\boldsymbol{X}_n = C_n'\boldsymbol{U}_n$ implies $(C_n')^{-1}\boldsymbol{X}_n = \boldsymbol{U}_n$ and hence $\boldsymbol{X}_n' C_n^{-1} = \boldsymbol{U}_n'$ and since

$$\Gamma_n = C_n' D_n C_n \text{ implies } \Gamma_n^{-1} = C_n^{-1} D_n^{-1} (C_n')^{-1},$$

it follows that

$$\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n = \boldsymbol{X}_n' C_n^{-1} D_n^{-1} (C_n')^{-1}\boldsymbol{X}_n = \boldsymbol{U}_n' D_n^{-1}\boldsymbol{U}_n,$$

i.e.,

$$\boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n = \sum_{j=1}^{n} \frac{U_j^2}{v_{j-1}}$$

(recall that $\boldsymbol{U}_n = [U_n, \ldots, U_1]'$ while D_n is a diagonal matrix with diagonal elements $v_{n-1}, \ldots, v_0$)

Maximum Likelihood Estimation: V

- since $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ & σ^2 determine Γ_n and since $v_j = r_j\sigma^2$, we can reexpress $-2\ln(L(\Gamma_n))$ as

$$\begin{aligned}
-2\ln\left(L(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)\right) &= n\ln(2\pi) + \ln(\det\Gamma_n) + \boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n \\
&= n\ln(2\pi) + \ln\left(\prod_{j=0}^{n-1} v_j\right) + \sum_{j=1}^{n}\frac{U_j^2}{v_{j-1}} \\
&= n\ln(2\pi\sigma^2) + \sum_{j=0}^{n-1}\ln(r_j) + \frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{U_j^2}{r_{j-1}} \\
&\stackrel{\text{def}}{=} n\ln(2\pi\sigma^2) + \sum_{j=0}^{n-1}\ln(r_j) + \frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{\sigma^2},
\end{aligned}$$

where we note that r_j's and $S(\boldsymbol{\phi},\boldsymbol{\theta})$ do not depend on σ^2

Maximum Likelihood Estimation: VI

- differentiating $-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)\big)$ with respect to σ^2 and setting resulting expression to zero yields MLE

$$\hat{\sigma}^2 = \frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{n} = \frac{1}{n}\sum_{j=1}^{n}\frac{U_j^2}{r_{j-1}}$$

- substituting $\hat{\sigma}^2$ for σ^2 in $-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2)\big)$ yields a so-called profile likelihood, which does not depend on σ^2

Maximum Likelihood Estimation: VII

- profile likelihood takes the form

$$
\begin{aligned}
-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta})\big) &= n\ln(2\pi\hat{\sigma}^2) + \sum_{j=0}^{n-1}\ln(r_j) + \frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{\hat{\sigma}^2} \\
&= n\ln(2\pi S(\boldsymbol{\phi},\boldsymbol{\theta})/n) + \sum_{j=0}^{n-1}\ln(r_j) + \frac{S(\boldsymbol{\phi},\boldsymbol{\theta})}{S(\boldsymbol{\phi},\boldsymbol{\theta})/n} \\
&= n + n\ln(2\pi/n) + n\ln(S(\boldsymbol{\phi},\boldsymbol{\theta})) + \sum_{j=0}^{n-1}\ln(r_j)
\end{aligned}
$$

Maximum Likelihood Estimation: VIII

- to evaluate $-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta})\big)$ for a particular ARMA(p,q) model, here are the steps we need to take

  1. compute ACVF for model out to lag $n-1$, setting $\sigma^2 = 1$
  2. get r_j's and compute 1-step-ahead predictions via recursions

$$\widehat{X}_{j+1} = \begin{cases} \sum_{k=1}^{j} \theta_{j,k} U_{j-k+1}, & 1 \le j < m; \\ \sum_{k=1}^{p} \phi_k X_{j-k+1} + \sum_{k=1}^{q} \theta_{j,k} U_{j-k+1}, & m \le j \le n-1 \end{cases}$$

     where r_j's & $\theta_{j,k}$'s come from IA applied to model ACVF (note: usually $v_j = r_j\sigma^2$, so setting $\sigma^2 = 1$ gives $r_j = v_j$); $U_1 = X_1$ and $U_j = X_j - \widehat{X}_j$ (as usual!); and $m = \max\{p,q\}$
  3. compute $S(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{j=1}^{n} U_j^2/r_{j-1}$ and $\sum_{j=0}^{n-1} \ln(r_j)$

- MLEs are settings $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ that minimize $-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta})\big)$
- side calculation gives corresponding $\hat{\sigma}^2 = S(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})/n$

Maximum Likelihood Estimation: IX

- can take alternative approach to evaluate $-2\ln\big(L(\boldsymbol{\phi})\big)$ in special case of AR(p) model

  1. use reverse L–D to generate r_j's and coefficients $\phi_{j,k}$ for best linear predictors of orders $j = p-1, \ldots, 1$, along with r_0
  2. compute 1-step-ahead predictions via recursions
  $$\widehat{X}_{j+1} = \begin{cases} \sum_{k=1}^{j} \phi_{j,k} X_{j-k+1}, & 1 \le j < p; \\ \sum_{k=1}^{p} \phi_k X_{j-k+1}, & p \le j \le n-1, \end{cases}$$
  along with innovations $U_{j+1} = X_{j+1} - \widehat{X}_{j+1}$ (recall that $U_1 = X_1$)
  3. compute $S(\boldsymbol{\phi}) = \sum_{j=1}^{p} U_j^2/r_{j-1} + \sum_{j=p+1}^{n} U_j^2$ & $\sum_{j=0}^{p-1} \ln(r_j)$

- MLEs are settings $\hat{\boldsymbol{\phi}}$ that minimize $-2\ln\big(L(\boldsymbol{\phi})\big)$
- side calculation gives corresponding $\hat{\sigma}^2 = S(\hat{\boldsymbol{\phi}})/n$

ML-Based Least Squares Estimation

- can argue that, for large n,

$$-2\ln\big(L(\boldsymbol{\phi},\boldsymbol{\theta})\big) = n + n\ln(2\pi/n) + n\ln(S(\boldsymbol{\phi},\boldsymbol{\theta})) + \sum_{j=0}^{n-1}\ln(r_j)$$

  depends mainly on $n\ln(S(\boldsymbol{\phi},\boldsymbol{\theta}))$, not $\sum_j \ln(r_j)$, in part because $r_j \to 1$ (for AR(p) models $r_j = 1$ for $j \geq p$)

- with $\sum_j \ln(r_j)$ dropped, minimization of

$$n + n\ln(2\pi/n) + n\ln(S(\boldsymbol{\phi},\boldsymbol{\theta}))$$

  is equivalent to minimization of $S(\boldsymbol{\phi},\boldsymbol{\theta})$, with minimizers $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\theta}}$ defining ML-based least squares estimators

- corresponding estimator for σ^2 is $\tilde{\sigma}^2 = S(\tilde{\boldsymbol{\phi}},\tilde{\boldsymbol{\theta}})/(n-p-q)$

- nonlinear optimization in general still required, but sometimes easier to find these LS estimates than the MLEs

Order Selection

- order selection can be based on AICC statistic:

$$\text{AICC} = -2\ln\left(L(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})\right) + \frac{2(p+q+1)n}{n-p-q-2},$$

  which, for a given model, is necessarily minimized by MLEs

- B&D, C&C and S&S discuss other order selection criteria (FPE for AR(p) models, AIC, BIC)
- see also Choi (1992), McQuarrie & Tsai (1998) and Stoica & Selén (2004)

Large Sample Distribution of MLEs: I

- consider causal & invertible ARMA(p,q) model

$$\phi(B)X_t = \theta(B)Z_t, \ \ \{Z_t\} \sim \text{WN}(0,\sigma^2)$$

- let $\{U_t\}$ and $\{V_t\}$ be AR(p) and AR(q) processes satisfying

$$\phi(B)U_t = Z_t \ \text{ and } \ \theta(B)V_t = Z_t,$$

and let $\boldsymbol{U}_t = [U_t, \ldots, U_{t-(p-1)}]'$ and $\boldsymbol{V}_t = [V_t, \ldots, V_{t-(q-1)}]'$

- letting $\boldsymbol{\beta} = [\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]'$ and letting $\hat{\boldsymbol{\beta}}$ denote the corresponding vector of MLEs, have, for large n,

$$\hat{\boldsymbol{\beta}} \approx \mathcal{N}(\boldsymbol{\beta}, V(\boldsymbol{\beta})/n),$$

where

$$V(\boldsymbol{\beta}) = \sigma^2 \begin{bmatrix} E\{\boldsymbol{U}_t\boldsymbol{U}_t'\} & E\{\boldsymbol{U}_t\boldsymbol{V}_t'\} \\ E\{\boldsymbol{V}_t\boldsymbol{U}_t'\} & E\{\boldsymbol{V}_t\boldsymbol{V}_t'\} \end{bmatrix}^{-1}$$

($q = 0$: $V(\boldsymbol{\beta}) = \sigma^2 [E\{\boldsymbol{U}_t\boldsymbol{U}_t'\}]^{-1}$; $p = 0$: $V(\boldsymbol{\beta}) = \sigma^2 [E\{\boldsymbol{V}_t\boldsymbol{V}_t'\}]^{-1}$)

Large Sample Distribution of MLEs: II

- easy to get $E\{\boldsymbol{U}_t\boldsymbol{U}_t'\}$ & $E\{\boldsymbol{V}_t\boldsymbol{V}_t'\}$; $E\{\boldsymbol{V}_t\boldsymbol{U}_t'\}$ is more work
- B&D give $V(\boldsymbol{\beta})$ for five special cases:

$$\text{AR(1):} \quad \begin{bmatrix} 1-\phi^2 \end{bmatrix}$$

$$\text{AR(2):} \quad \begin{bmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{bmatrix}$$

$$\text{MA(1):} \quad \begin{bmatrix} 1-\theta^2 \end{bmatrix}$$

$$\text{MA(2):} \quad \begin{bmatrix} 1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2 \end{bmatrix}$$

$$\text{ARMA(1,1):} \quad \frac{1+\phi\theta}{(\phi+\theta)^2}\begin{bmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\phi^2)(1-\theta^2) \\ -(1-\phi^2)(1-\theta^2) & (1-\theta^2)(1+\phi\theta) \end{bmatrix}$$

- can plug in estimates $\hat{\phi}_i$ & $\hat{\theta}_j$ (OK for large n & small $p+q$)

Example – Atomic Clock Series: VI

model	suggested by	fitted by	$\hat{\phi}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	AICC
MA(1)	ACF CIs	MM	—	-0.44	—	—	6108.42
		IA	—	-0.59	—	—	6067.05
		ML	—	-0.73	—	—	6051.33
MA(2)	IA CIs	IA	—	-0.59	-0.13	—	6024.04
		ML	—	-0.61	-0.18	—	6016.66
MA(3)	IA AICC	IA	—	-0.59	-0.13	-0.07	6013.40
		ML	—	-0.60	-0.14	-0.08	6012.65
ARMA(1,1)	ACF/PACF	H–R	0.27	-0.87	—	—	6011.53
		ML	0.27	-0.86	—	—	**6011.52**

- ML AICCs for
  - MA(4), ARMA(1,2) & ARMA(2,1): 6014.65, 6013.20, 6013.14
  - AR(1), AR(2) & AR(3): 6190.78, 6136.49, 6095.82
  - AR(9), AR(10) & AR(11): 6033.24, 6031.08, 6031.73

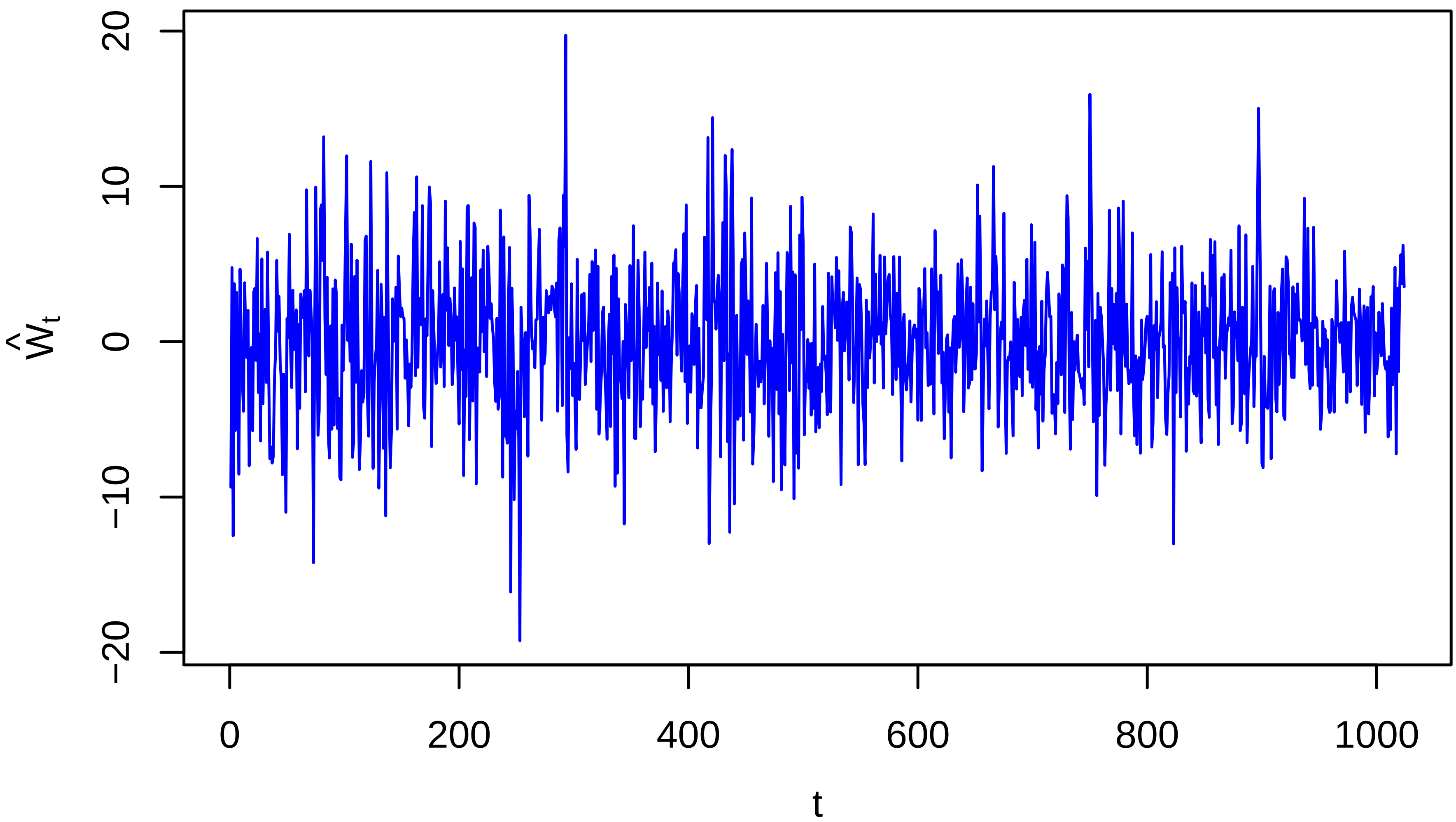
Diagnostic Checking

- diagnostic tests based on normalized innovations corresponding to fitted model (will refer to these as 'residuals' for convenience):

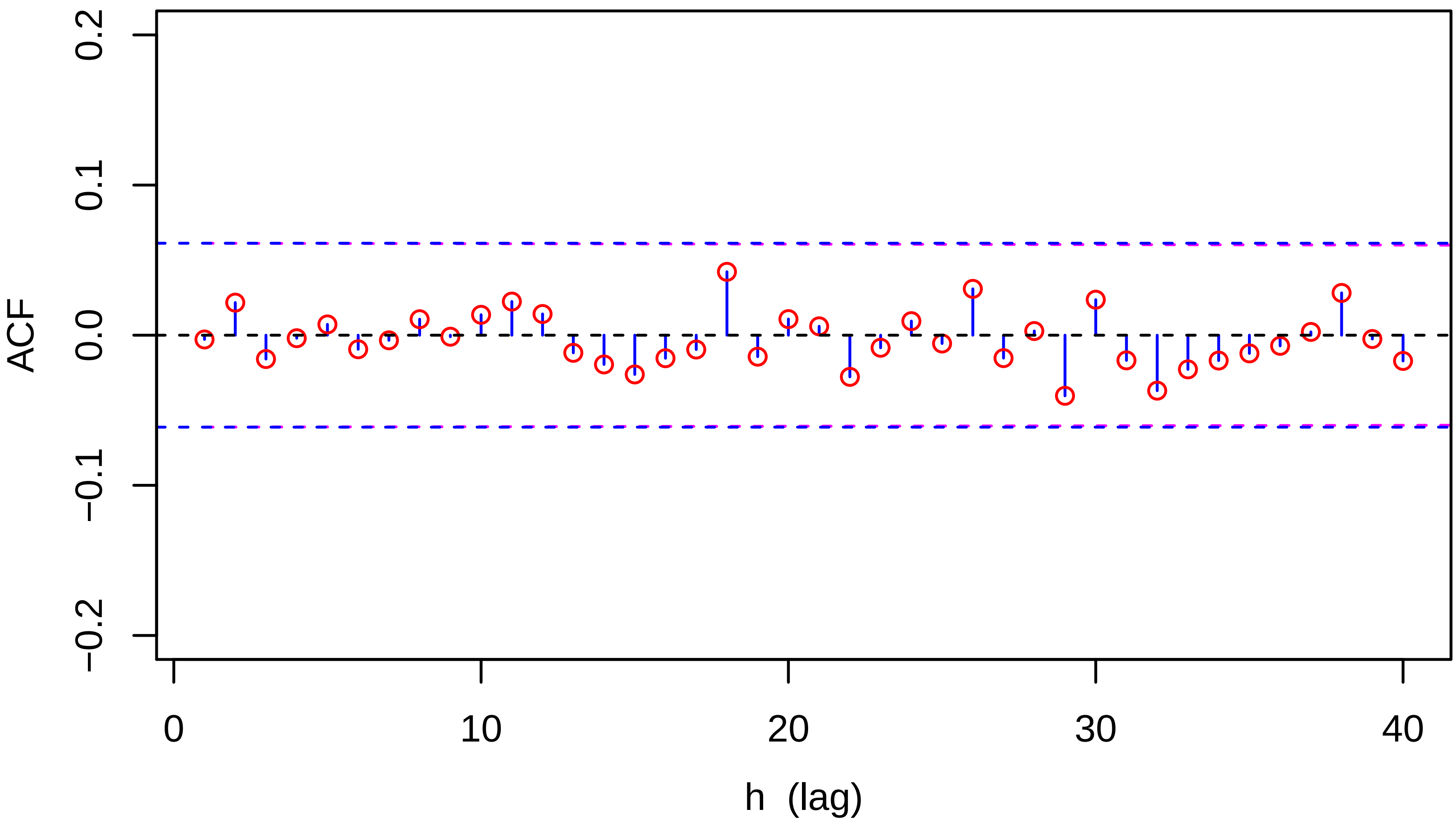
$$\widehat{W}_t = \frac{U_t}{\sqrt{r_{t-1}}} = \frac{X_t - \widehat{X}_t}{\sqrt{r_{t-1}}}, \quad t = 1, \ldots, n$$

- if fitted model is good representation for time series, $\widehat{W}_t$'s should resemble zero-mean white noise (but won't be exactly so)
- tests include
  - informal assessment based on plot of $\widehat{W}_t$'s
  - sample ACF and PACF for $\widehat{W}_t$'s
  - checks on hypothesis of randomness
- let's illustrate diagnostic checking by looking at fitted ARMA(1,1) model for atomic clock series

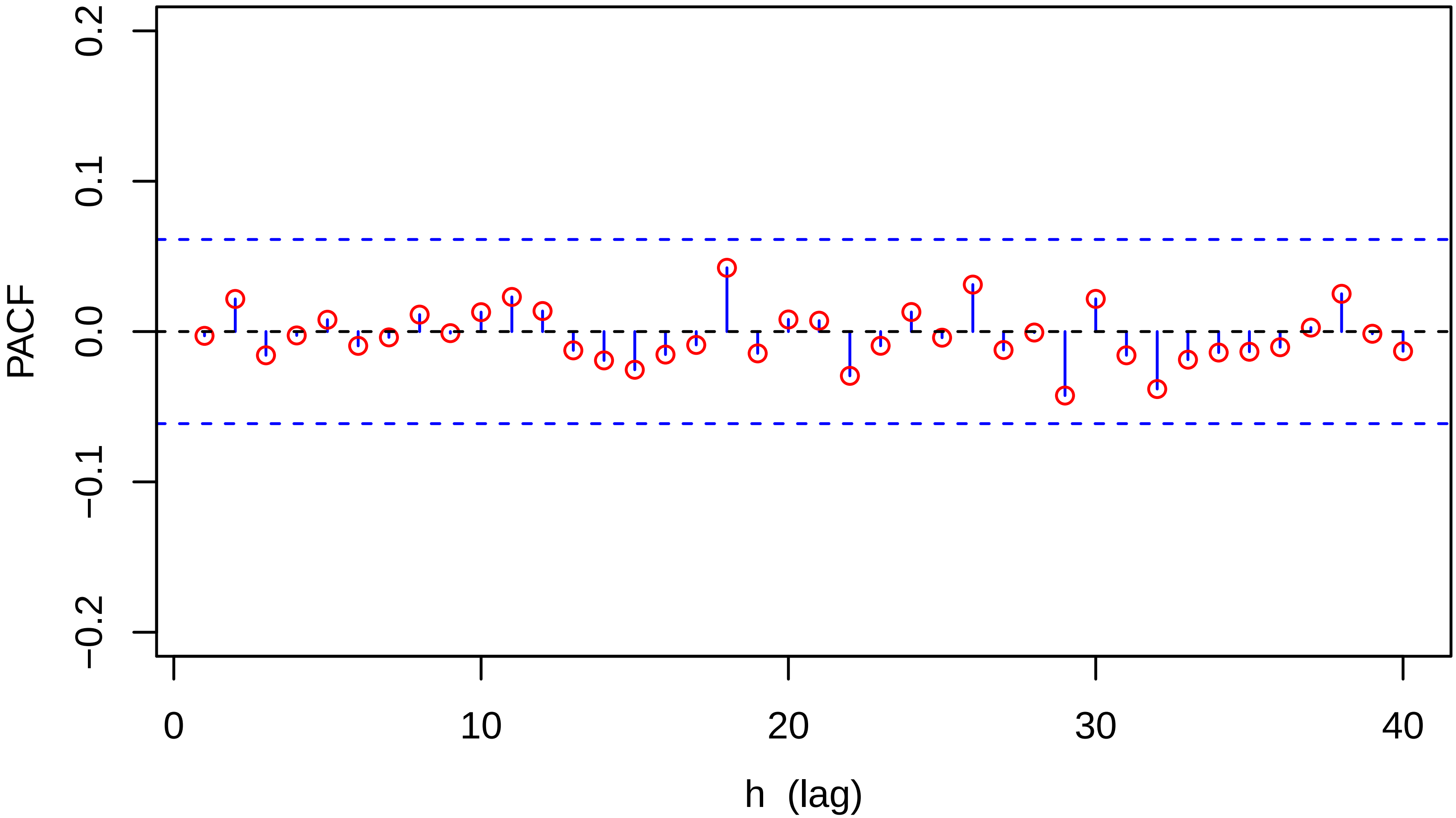
ARMA(1,1) Residuals $\widehat{W}_t$ for Atomic Clock



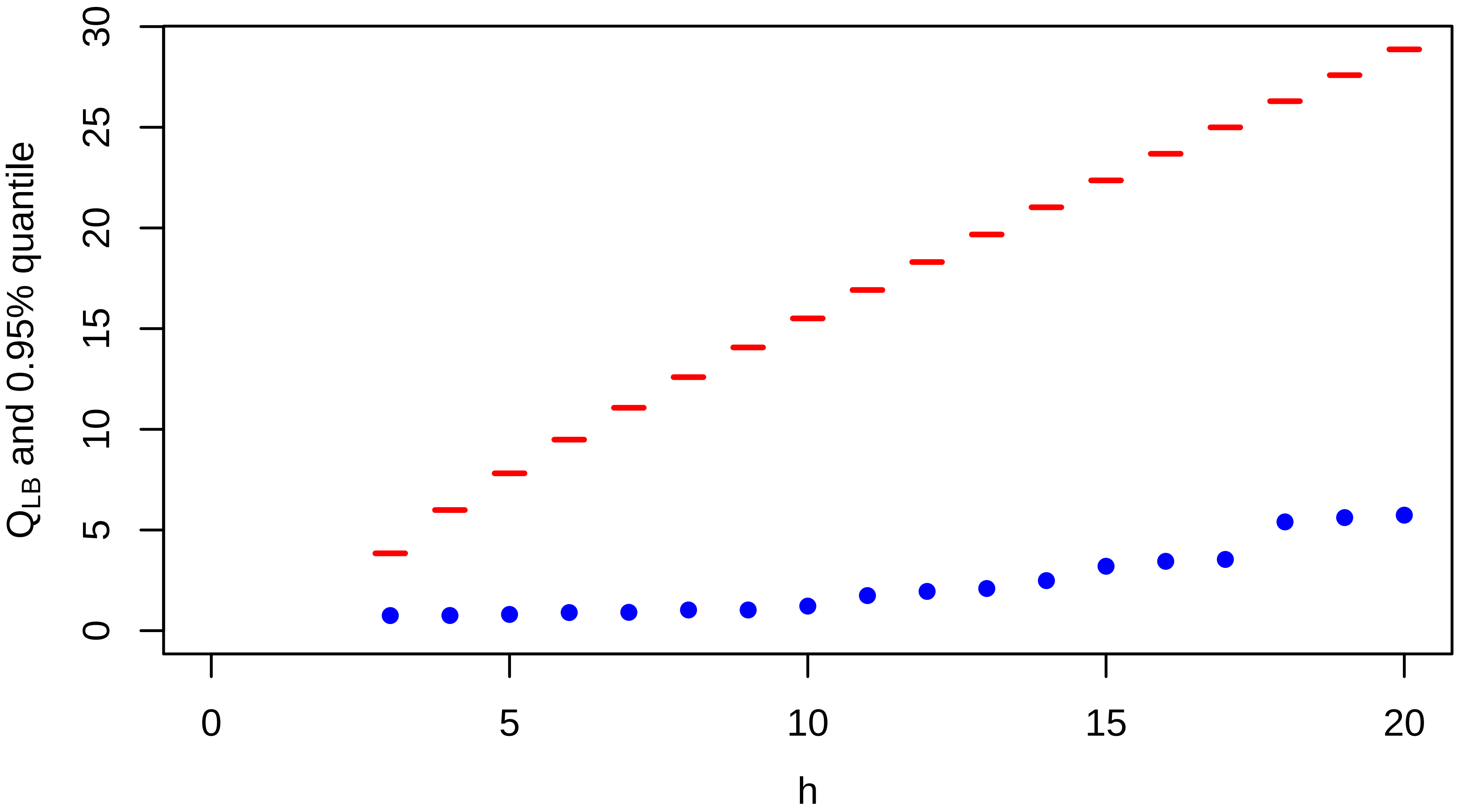
Sample ACF for ARMA(1,1) Residuals $\widehat{W}_t$



Sample PACF for ARMA(1,1) Residuals $\widehat{W}_t$



Portmanteau Tests of ARMA(1,1) Residuals $\widehat{W}_t$

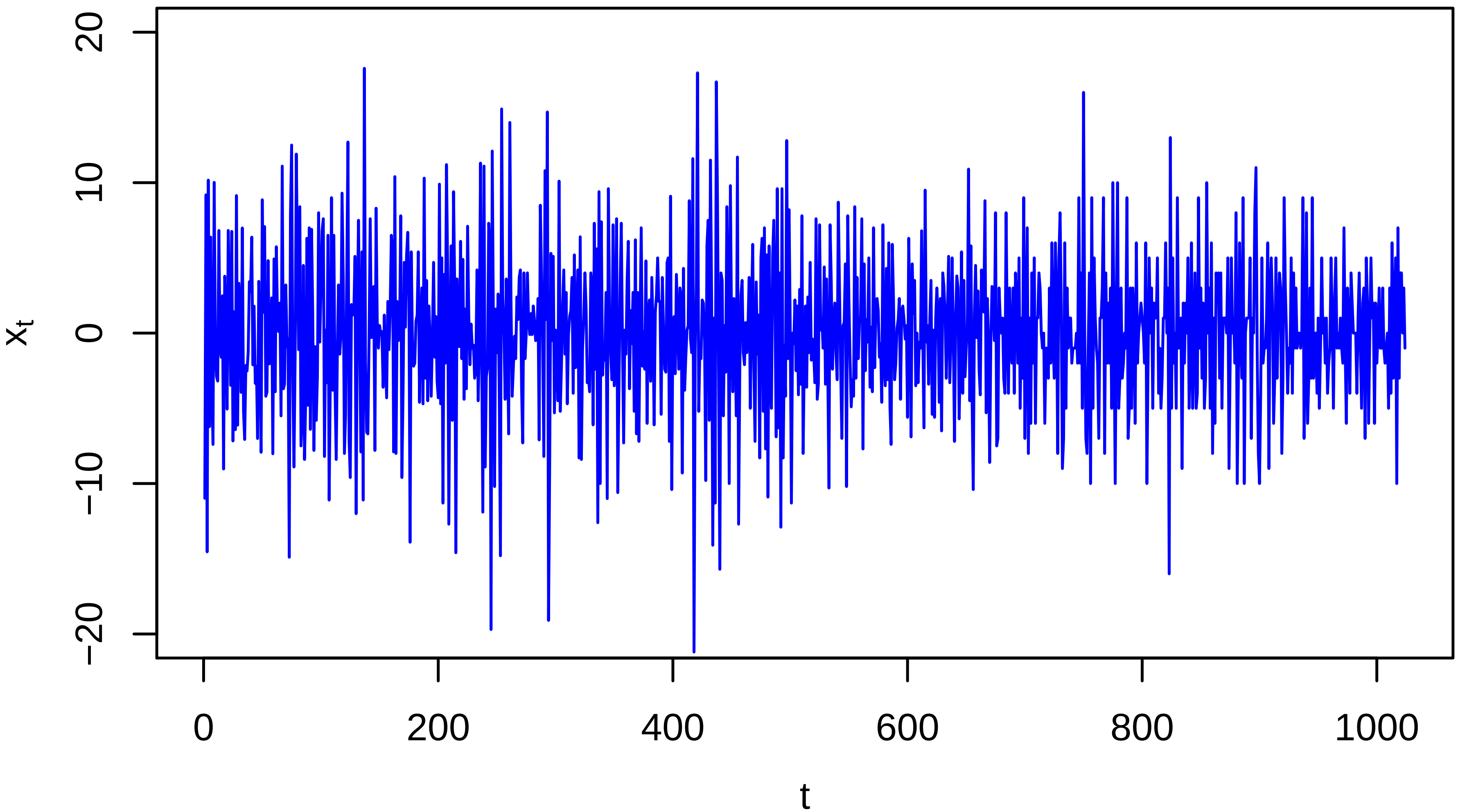


Example – Atomic Clock Series: VII

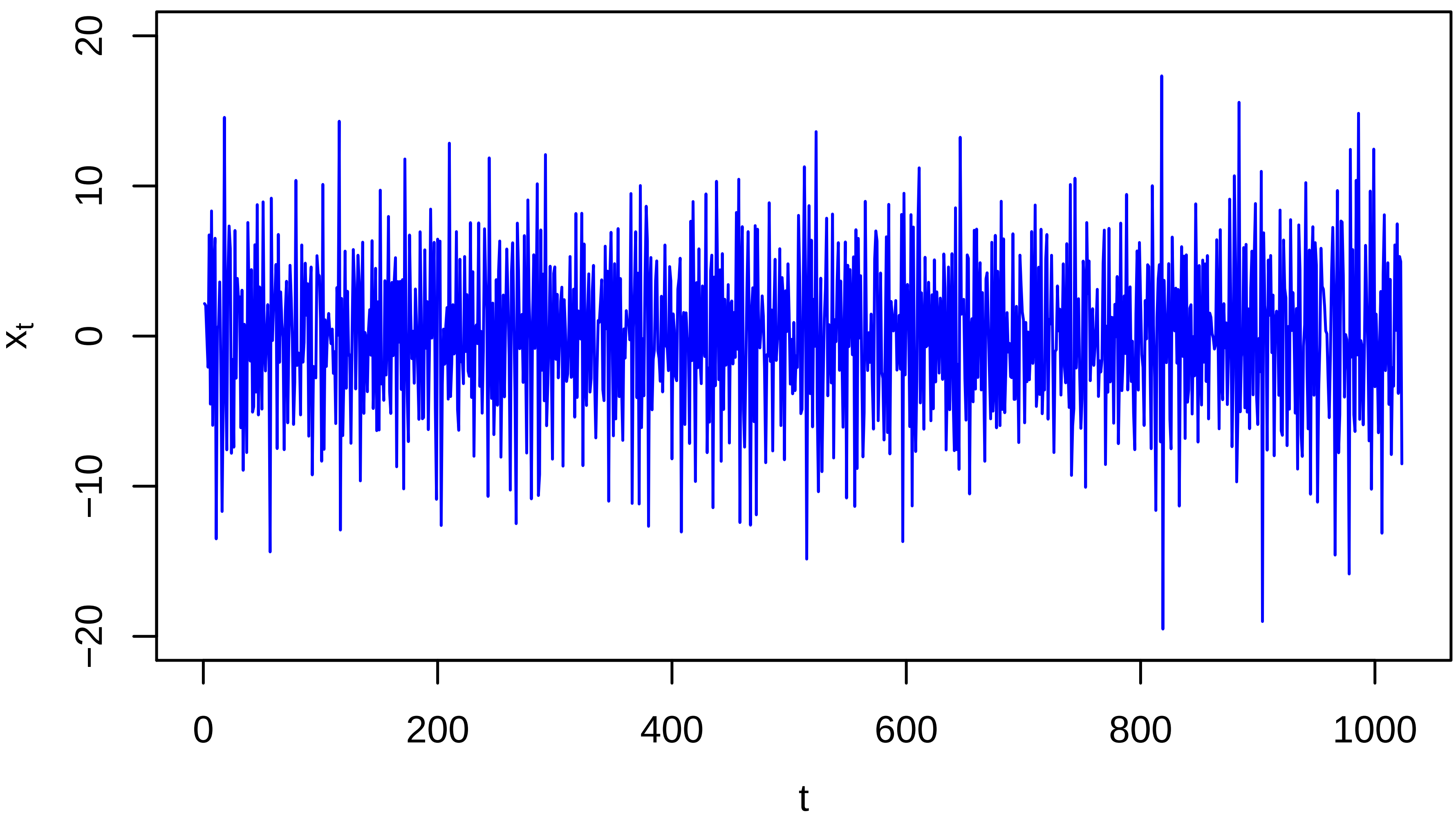
test	expected value	test statistic	p-value
turning point	681.3	709	**0.040**
difference-sign	511.5	515	0.705
rank	261888	262439	0.920
runs	513.0	506	0.662

AR method	AICC order	AIC order
Yule–Walker	0	0
Burg	0	0
OLS	0	**3**
MLE	0	0

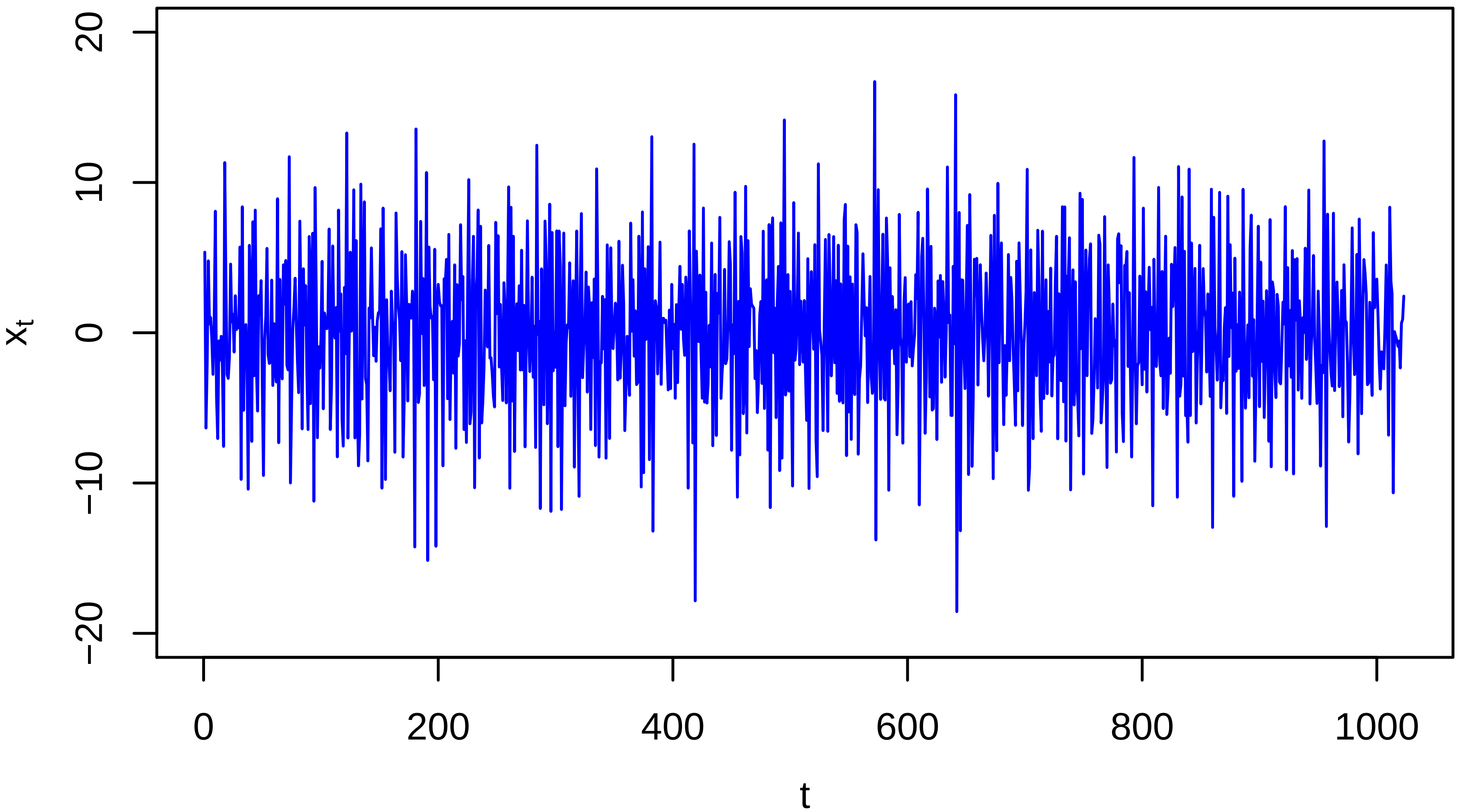
Atomic Clock Time Series



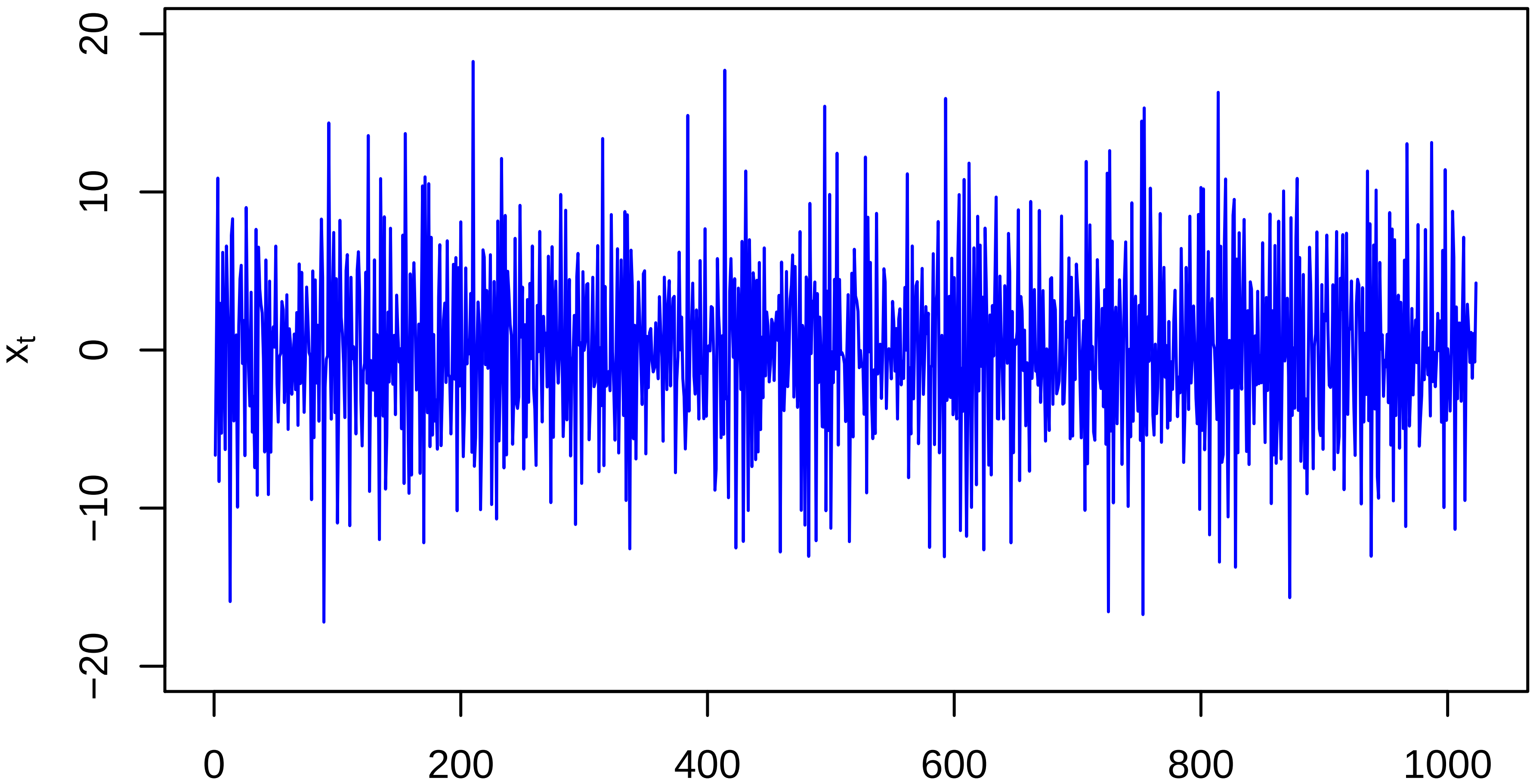
Simulated Atomic Clock Time Series: I



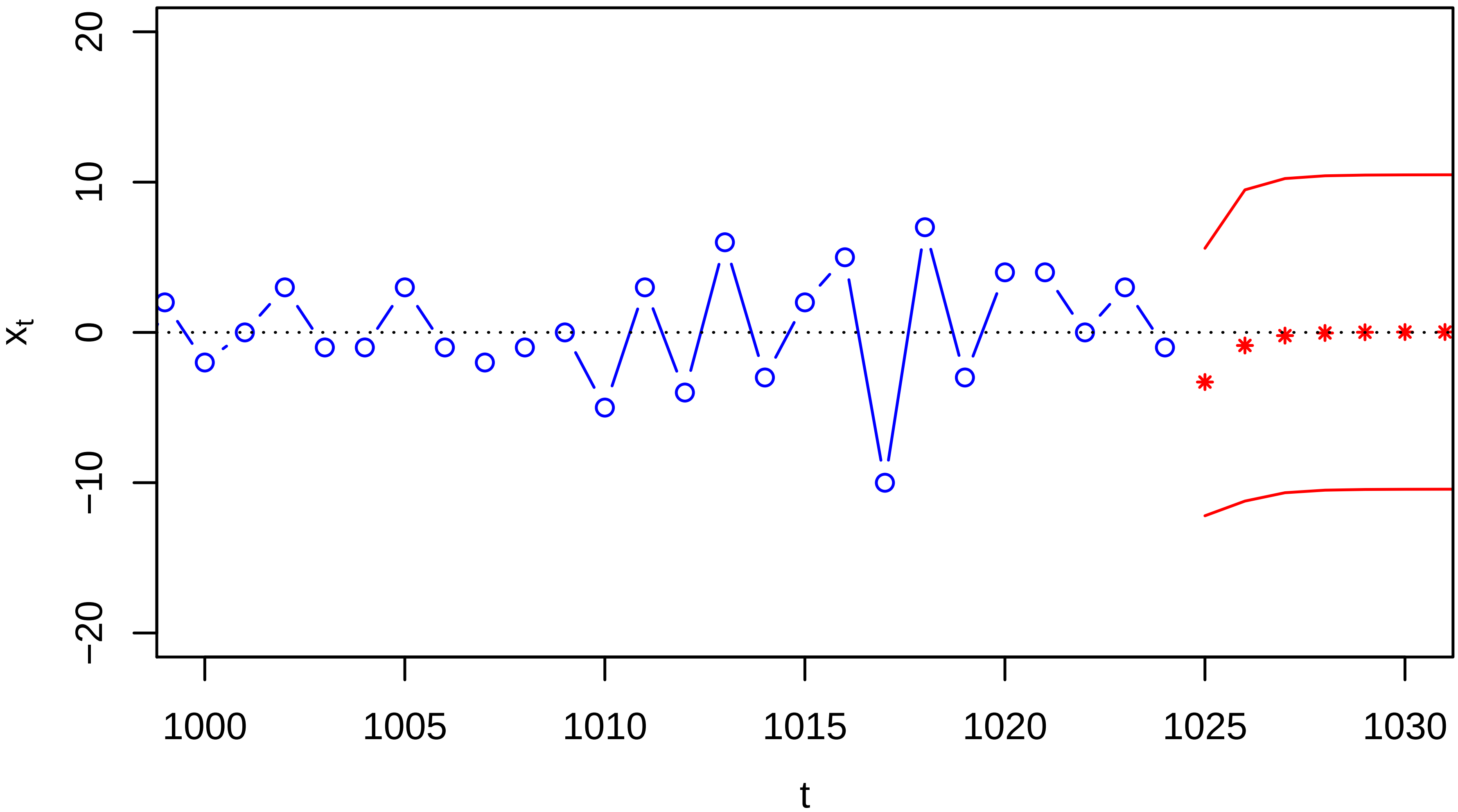
Simulated Atomic Clock Time Series: II



Simulated Atomic Clock Time Series: III



Forecasting Atomic Clock Time Series



References

- B. Choi (1992), *ARMA Model Identification*, New York: Springer-Verlag
- A. D. R. McQuarrie and C.-L. Tsai (1998), *Regression and Time Series Model Selection*, Singapore: World Scientific
- P. Stoica and Y. Selén (2004), 'Model-Order Selection: A Review of Information Criterion Rules,' *IEEE Signal Processing Magazine*, **21**, pp. 36–47