Differencing Revisited: I

- ARIMA(p,d,q) processes predicated on notion of dth order differencing of a time series $\{X_t\}$: for $d = 1$ and 2, have

$$\begin{aligned}\nabla X_t &\stackrel{\text{def}}{=} (1-B)X_t = X_t - X_{t-1}\\ \nabla^2 X_t &\stackrel{\text{def}}{=} \nabla(\nabla X_t) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\\ &= (1-B)^2 X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

- in general

$$\nabla^d X_t \stackrel{\text{def}}{=} (1-B)^d X_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}$$

- differencing a process is a severe hammering
  - changes a random walk into white noise
  - changes a stationary process into another stationary process, but with substantially altered covariance structure

Differencing Revisited: II

- motivation for *fractional differencing* is to define an operation that is not as drastic as backward differencing
- leads to an interesting class of processes known as *fractionally differenced processes* (introduced by Granger & Joyeux, 1980, and Hosking, 1981)
- recalling that
  - white noise is an ARIMA$(0, 0, 0)$ process and
  - random walk is an ARIMA$(0, 1, 0)$ process,

  allows us to define ARIMA$(0, d, 0)$ processes with $0 < d < 1$ that offer compromises between two extreme cases
- leads to definition of ARIMA(p, d, q) process for which d is not limited to nonnegative integers

Fractionally Differenced Processes: I

- dth order difference of $Z_t \sim \text{WN}(0, \sigma^2)$ is, for $d = 0, 1, 2, \ldots$,

$$X_t = \nabla^d Z_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k Z_{t-k}$$

$$= \sum_{k=0}^{d} \frac{d!}{k!(d-k)!} (-1)^k Z_{t-k}$$

$$= \sum_{k=0}^{d} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k Z_{t-k}$$

$$= \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k Z_{t-k}$$

since $1/\Gamma(j) = 0$ for $j = 0, -1, -2, \ldots$

Fractionally Differenced Processes: II

- to conform to convention in literature, let $\delta \stackrel{\text{def}}{=} -d$ so that
$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k Z_{t-k}$$
for $\delta = 0, -1, -2, \ldots$
- above infinite sum also makes sense for $-1/2 \leq \delta < 1/2$:
$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k},$$
where
$$\psi_k \stackrel{\text{def}}{=} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k = \frac{\Gamma(k+\delta)}{\Gamma(k+1)\Gamma(\delta)}$$
(unless $\delta = 0$, $\psi_k \neq 0$ for $k \geq 1$, so have true infinite sum)
- defines fractionally differenced (FD) process with parameter δ

Fractionally Differenced Processes: III

- using $\Gamma(x) = (x-1)\Gamma(x-1)$, can write

$$\psi_k = \frac{\Gamma(k+\delta)}{\Gamma(k+1)\Gamma(\delta)} = \frac{(k+\delta-1)\Gamma(k+\delta-1)}{k\Gamma(k)\Gamma(\delta)} = \frac{k+\delta-1}{k}\psi_{k-1}$$

  for $k = 1, 2, \ldots$ (starting with $\psi_0 = 1$)

- FD process $\{X_t\}$ is stationary with ACVF

$$\gamma(h) = \sigma^2 \frac{\sin(\pi\delta)\Gamma(1-2\delta)\Gamma(h+\delta)}{\pi\Gamma(h-\delta+1)}$$

- can reexpress this in a computationally friendly manner:

$$\gamma(0) = \sigma^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} \text{ and } \gamma(h) = \gamma(h-1)\frac{h+\delta-1}{h-\delta}, \quad h = 1, 2, \ldots$$

- implies that ACF at unit lag is

$$\rho(1) = \frac{\delta}{1-\delta} \text{ and hence } -\tfrac{1}{3} \le \rho(1) < 1 \text{ since } -\tfrac{1}{2} \le \delta < \tfrac{1}{2}$$

Fractionally Differenced Processes: IV

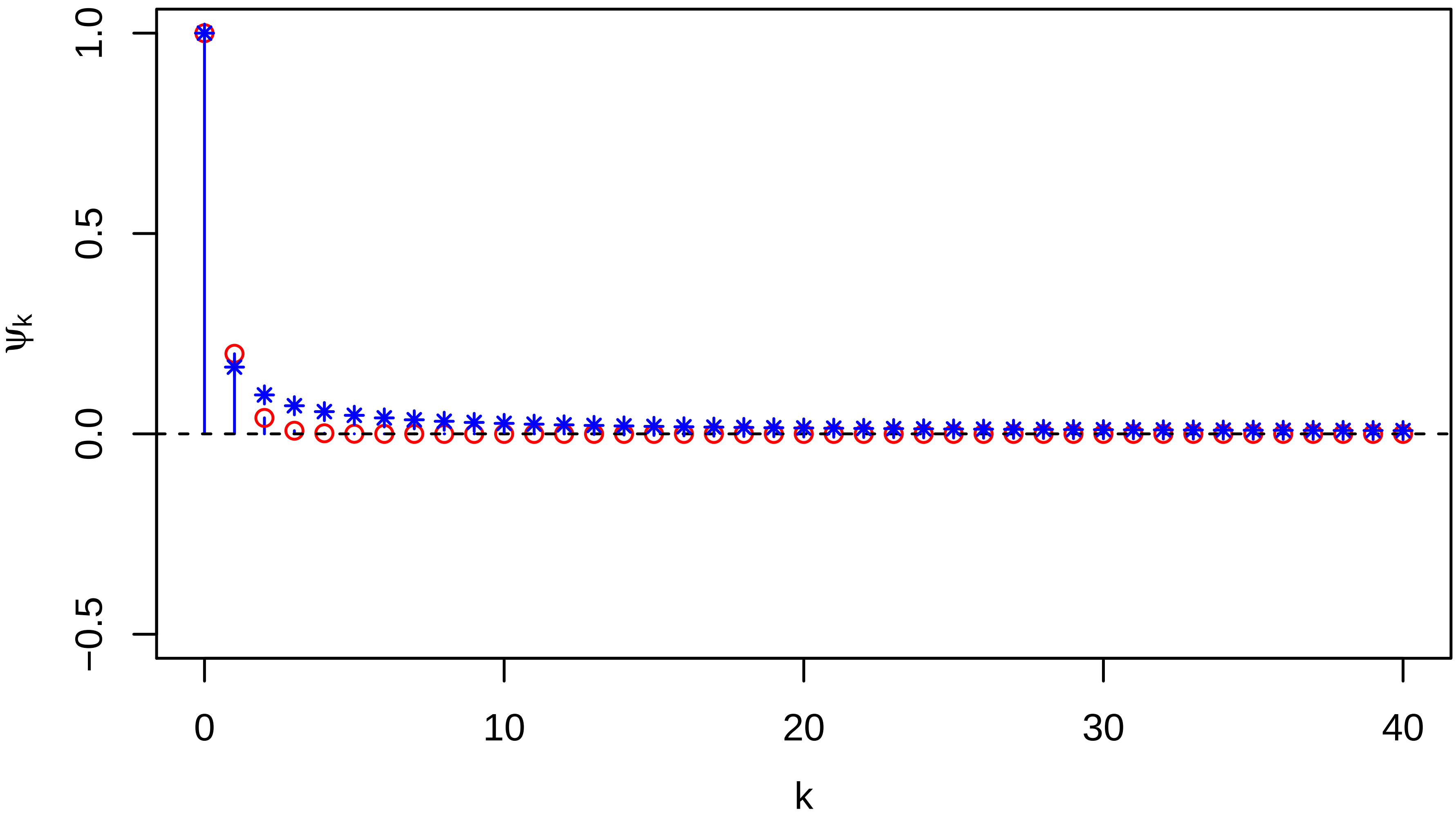
- PACF for FD process has a remarkably simple form:

$$\phi_{h,h} = \frac{\delta}{h - \delta}, \quad h = 1, 2, \ldots$$

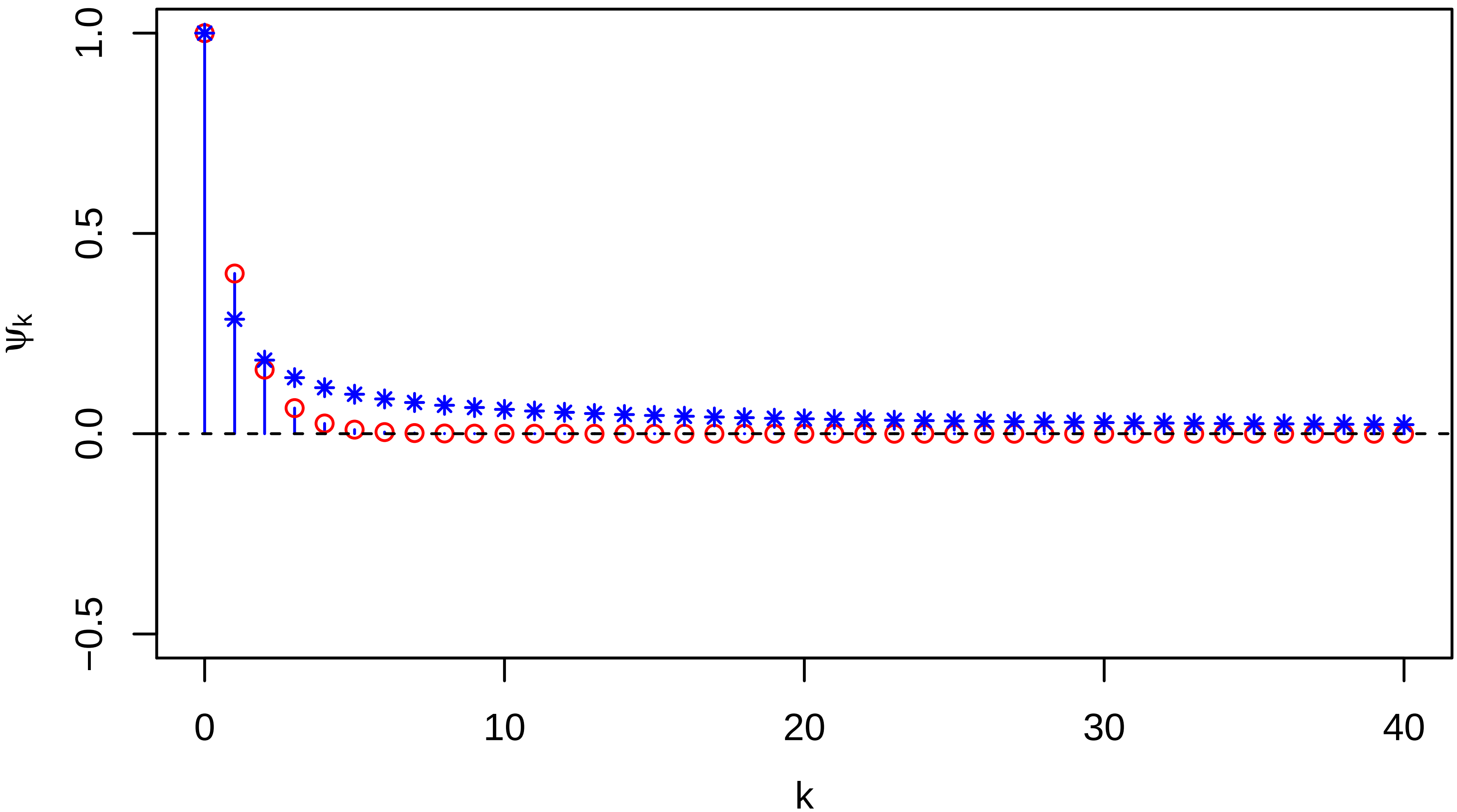
  (will be put to good use later on)

- following overheads show ψ_k's, ACFs and PACFs for AR(1) and FD processes with parameters ϕ and δ adjusted so that $\rho(1)$'s are identical (and hence also $\phi_{1,1}$)

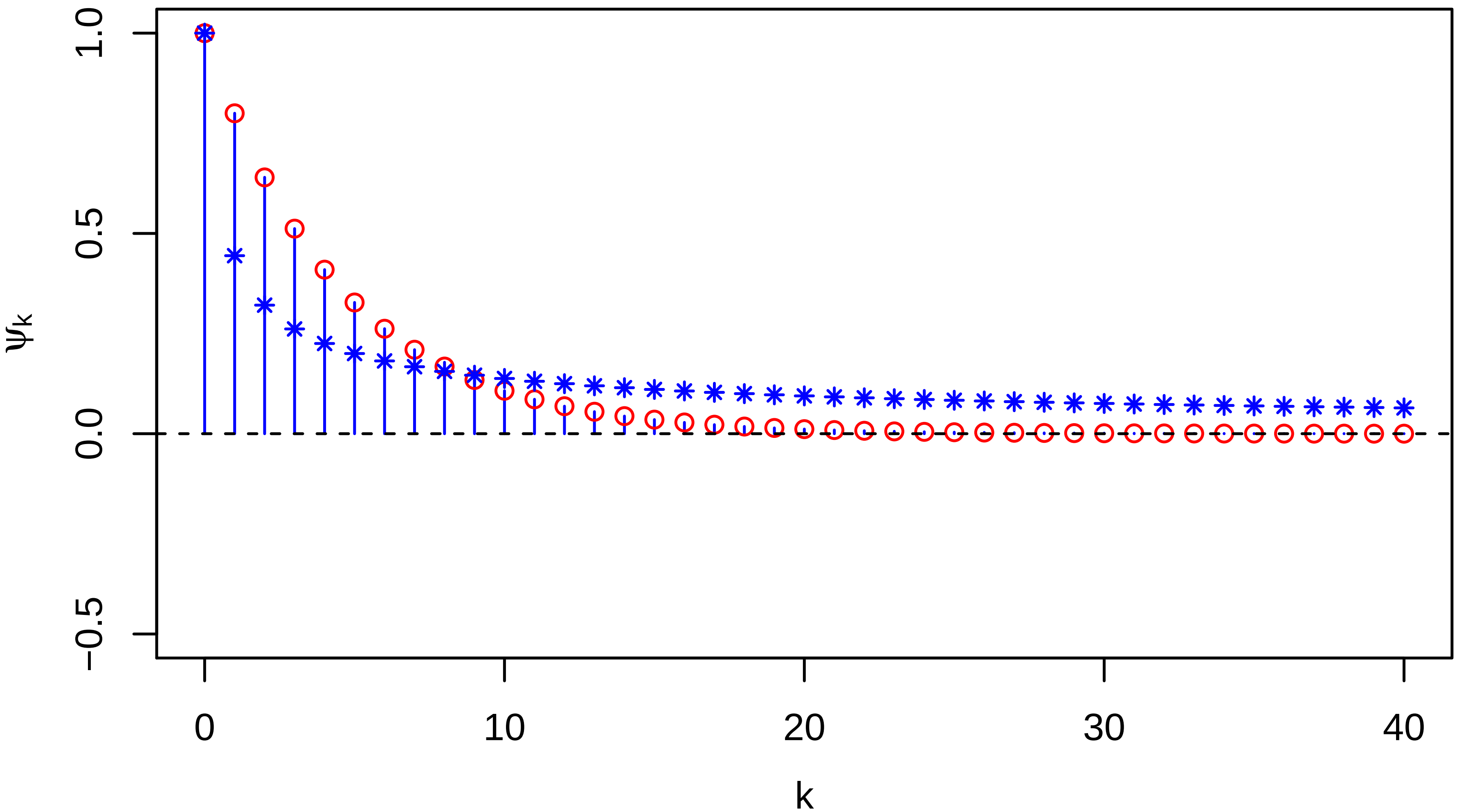
Comparison of AR (∘) and FD (∗) ψ_k's, $\rho(1) = 0.2$



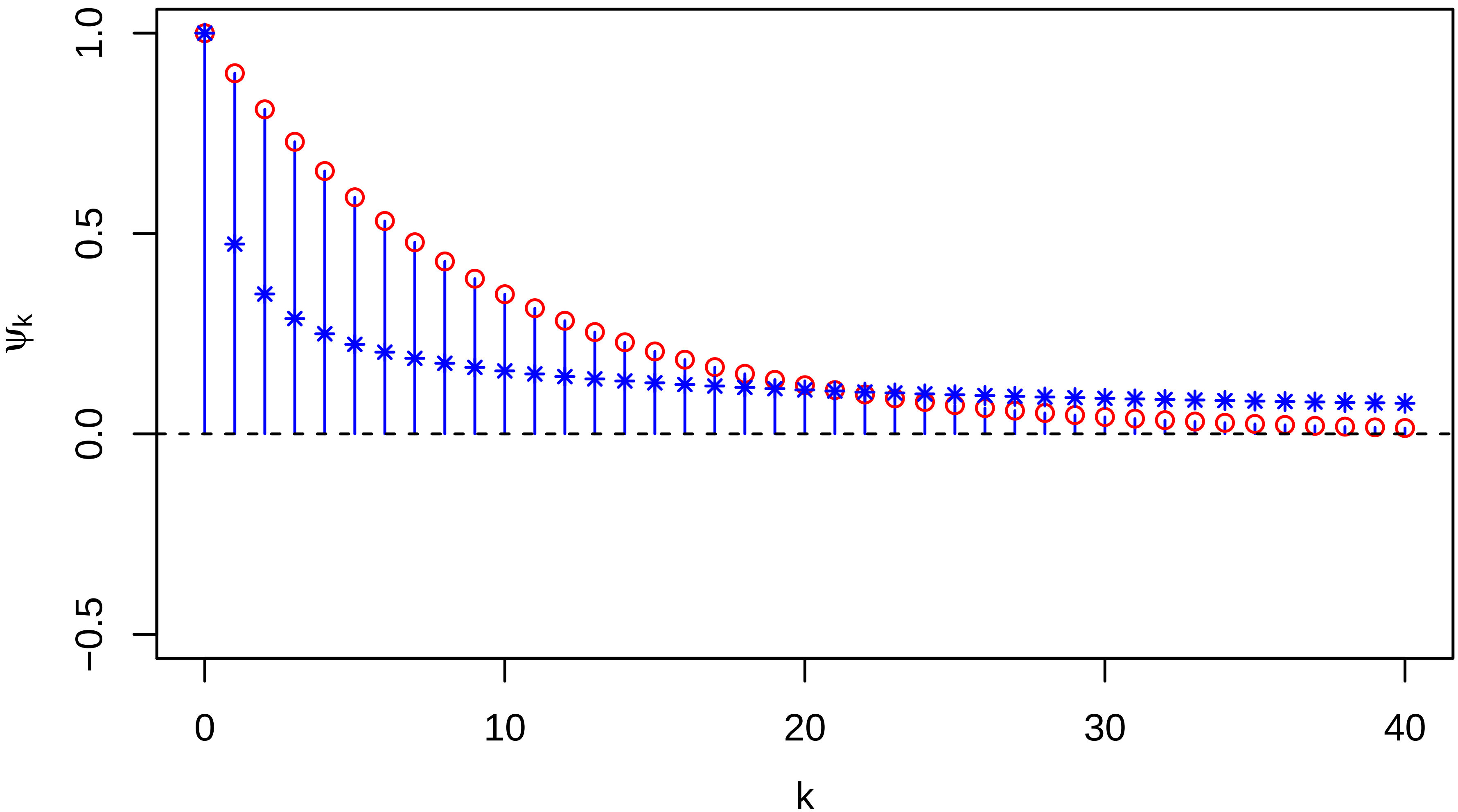
Comparison of AR (∘) and FD (∗) ψ_k's, $\rho(1) = 0.4$



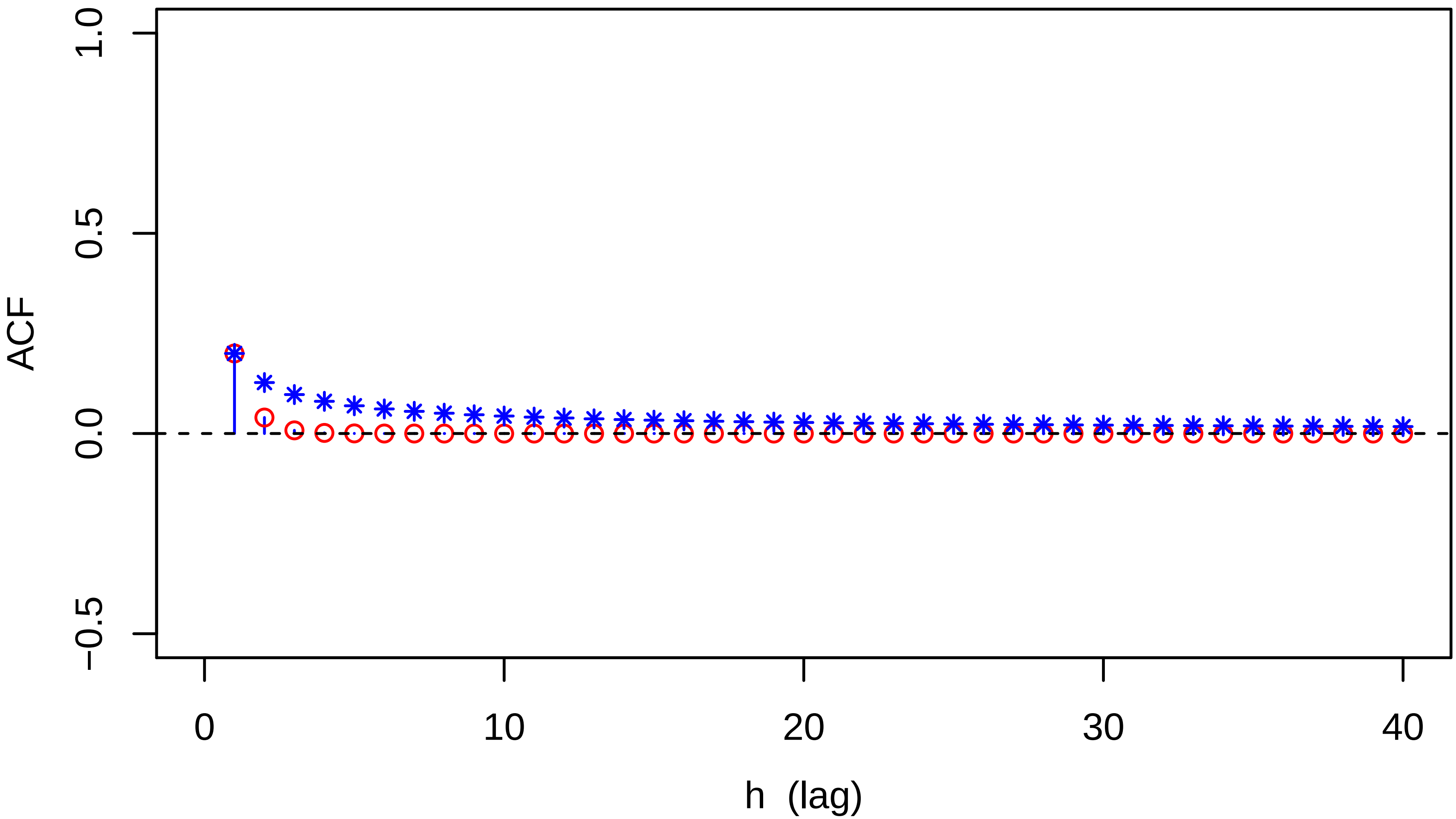
Comparison of AR (∘) and FD (∗) ψ_k's, $\rho(1) = 0.8$



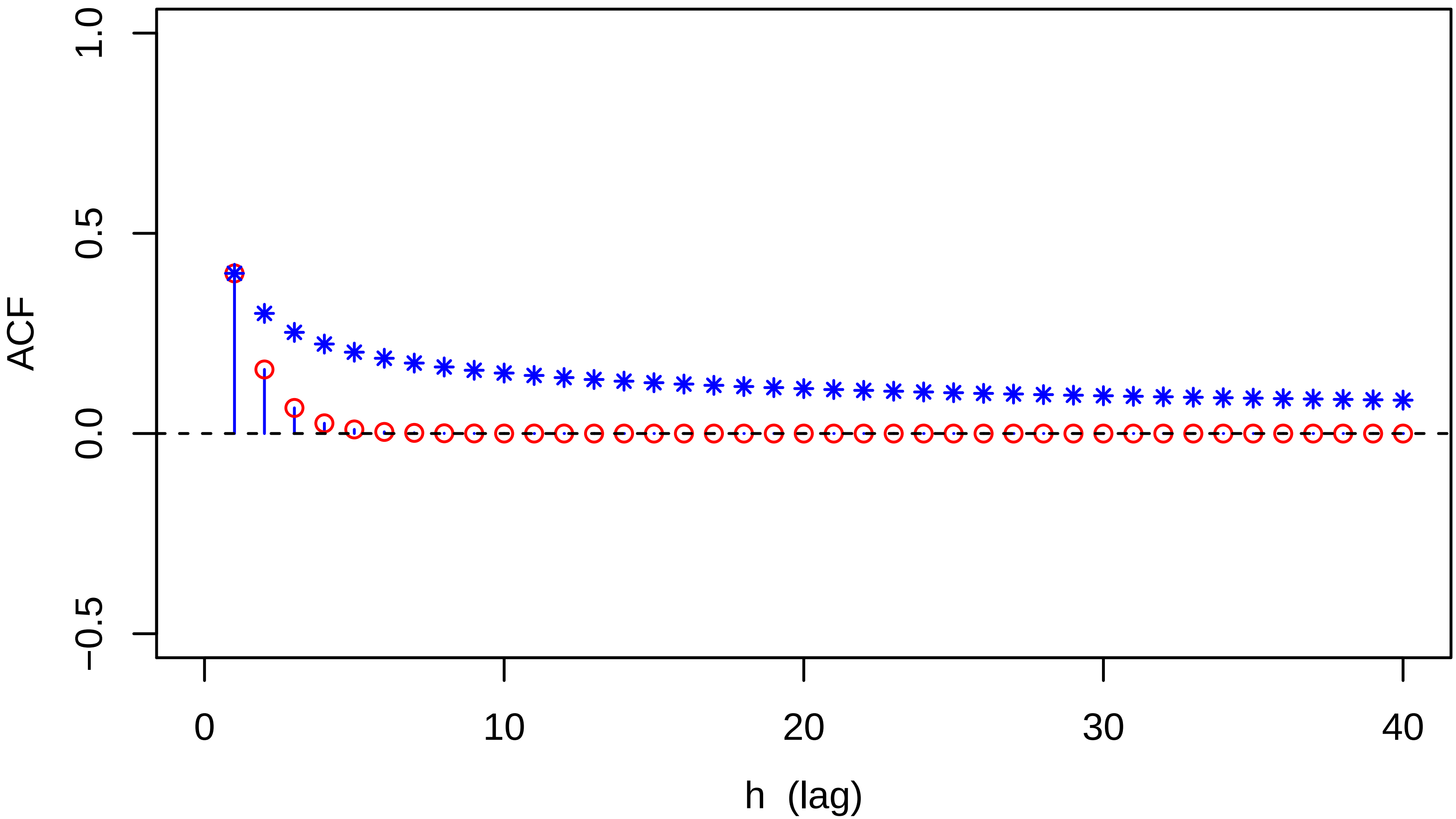
Comparison of AR (∘) and FD (∗) ψ_k's, $\rho(1) = 0.9$



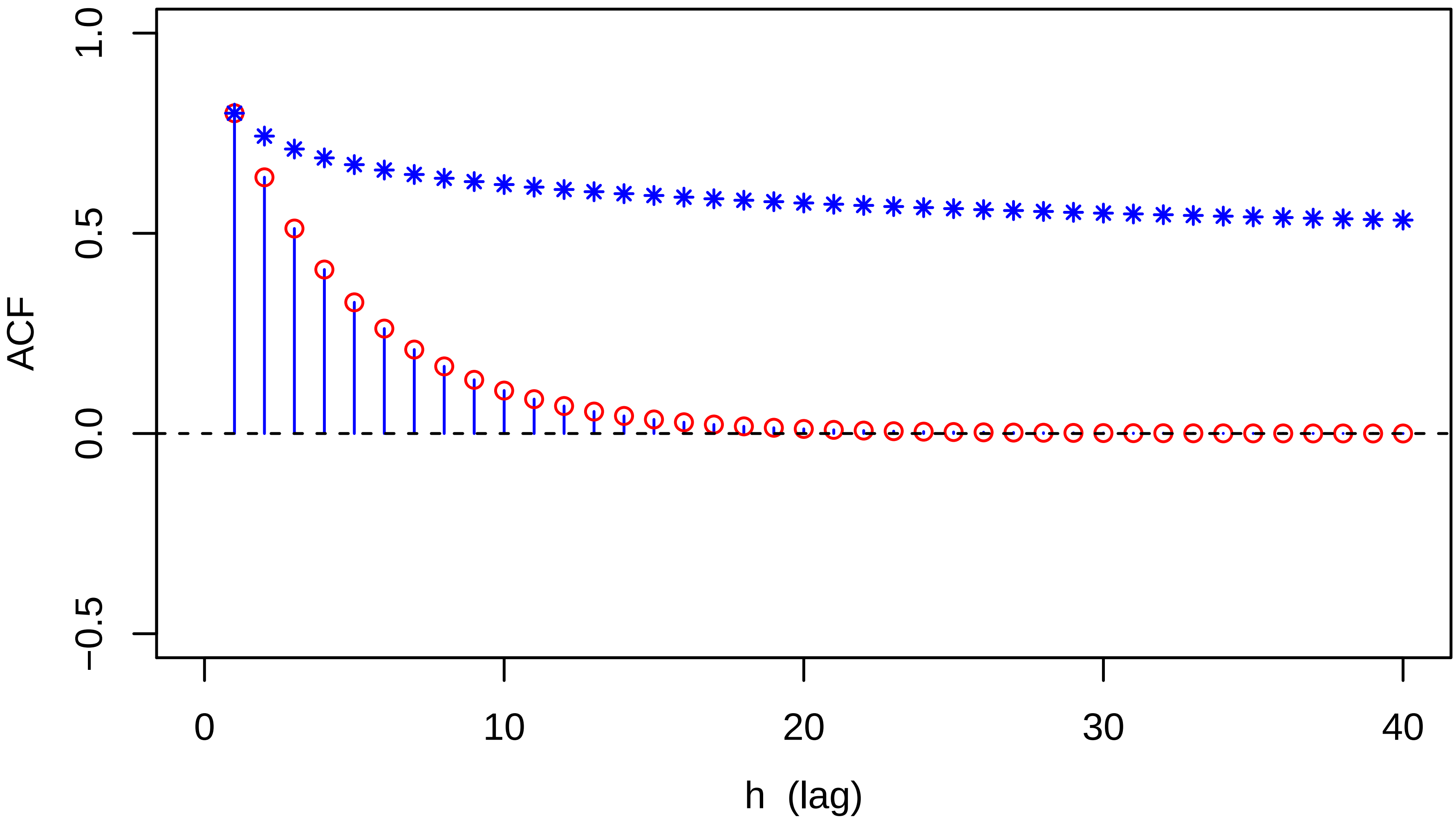
Comparison of AR (∘) and FD (∗) ACFs, $\rho(1) = 0.2$

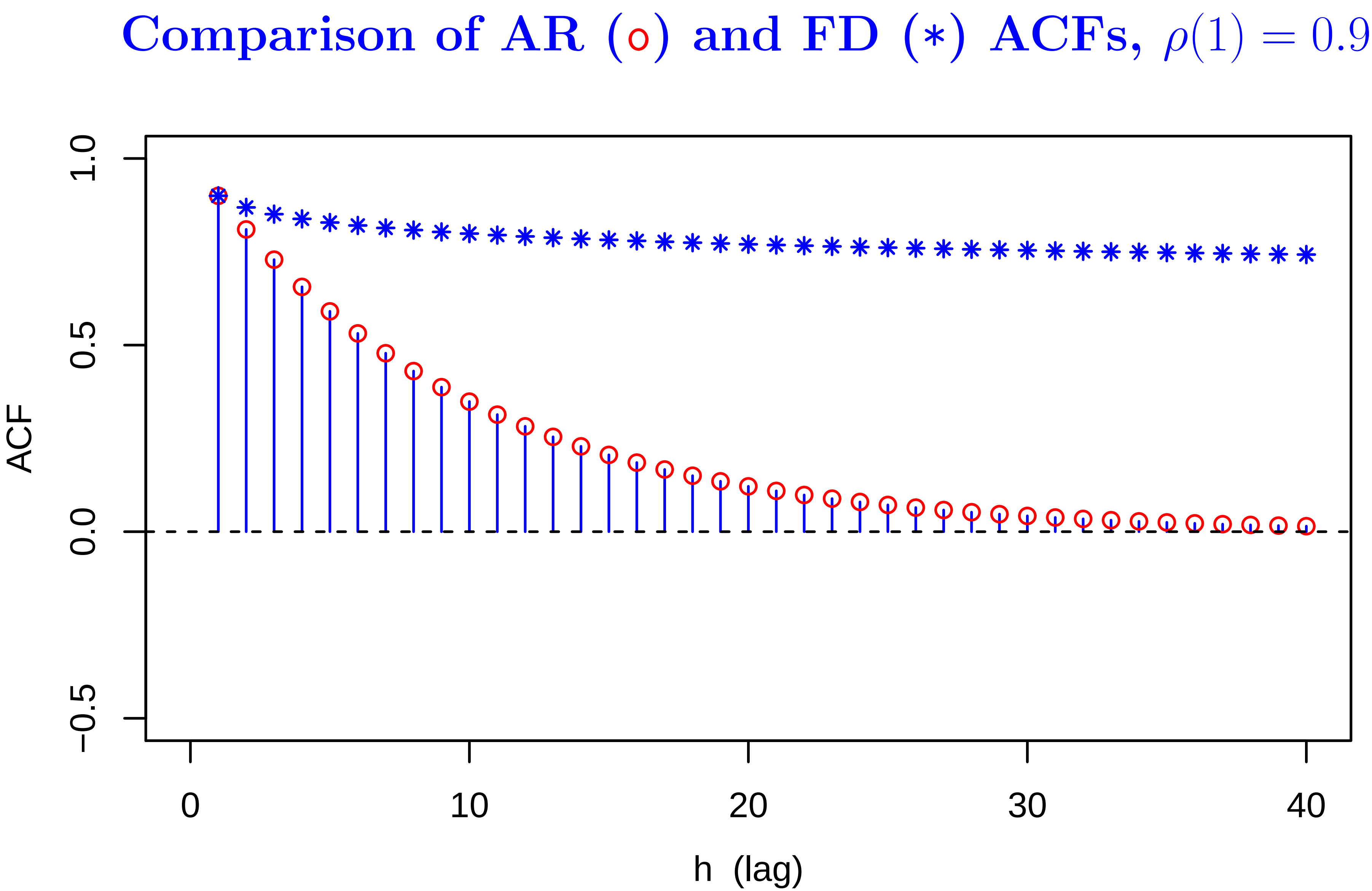


Comparison of AR (∘) and FD (∗) ACFs, $\rho(1) = 0.4$

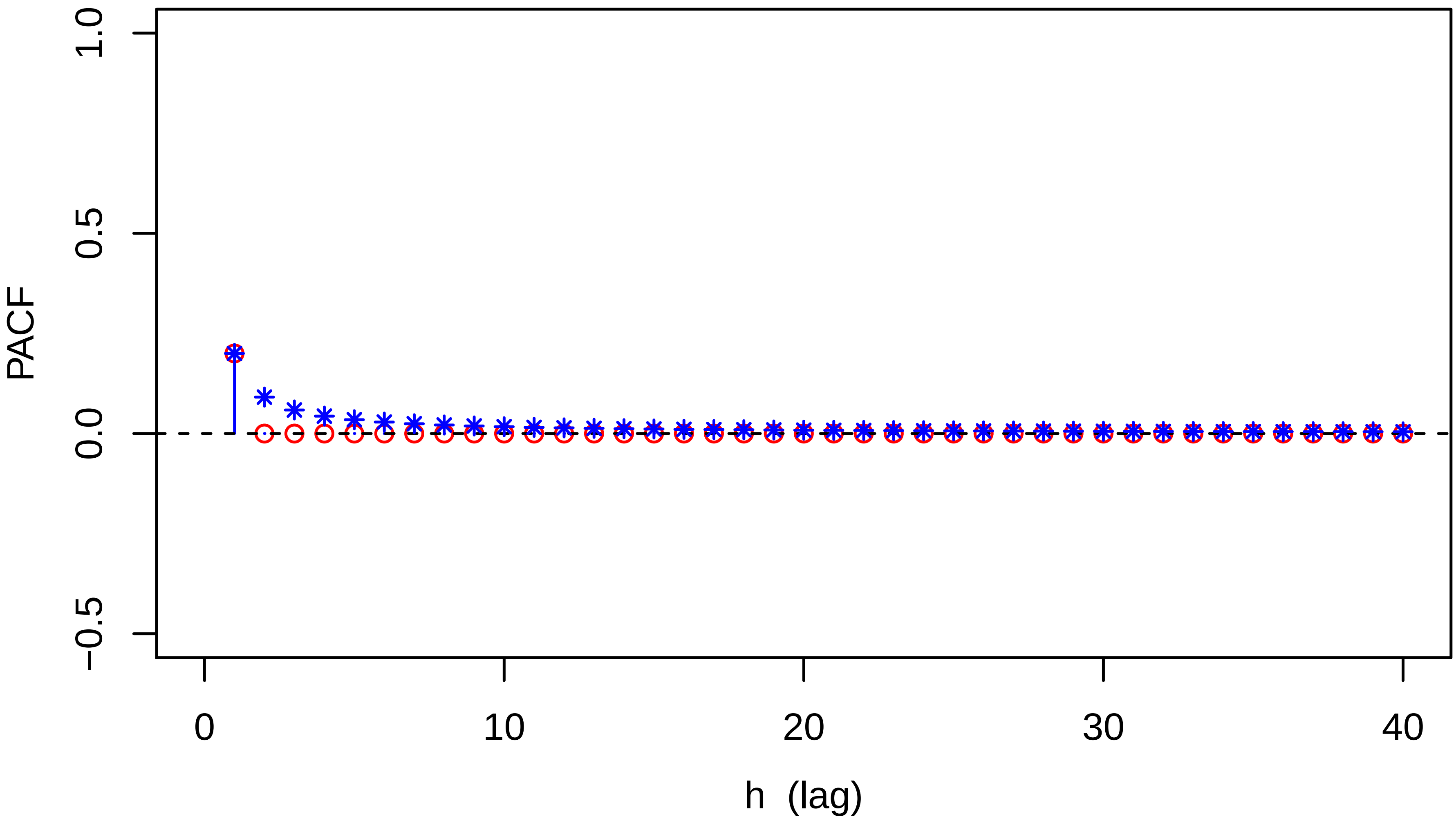


Comparison of AR (∘) and FD (∗) ACFs, $\rho(1) = 0.8$

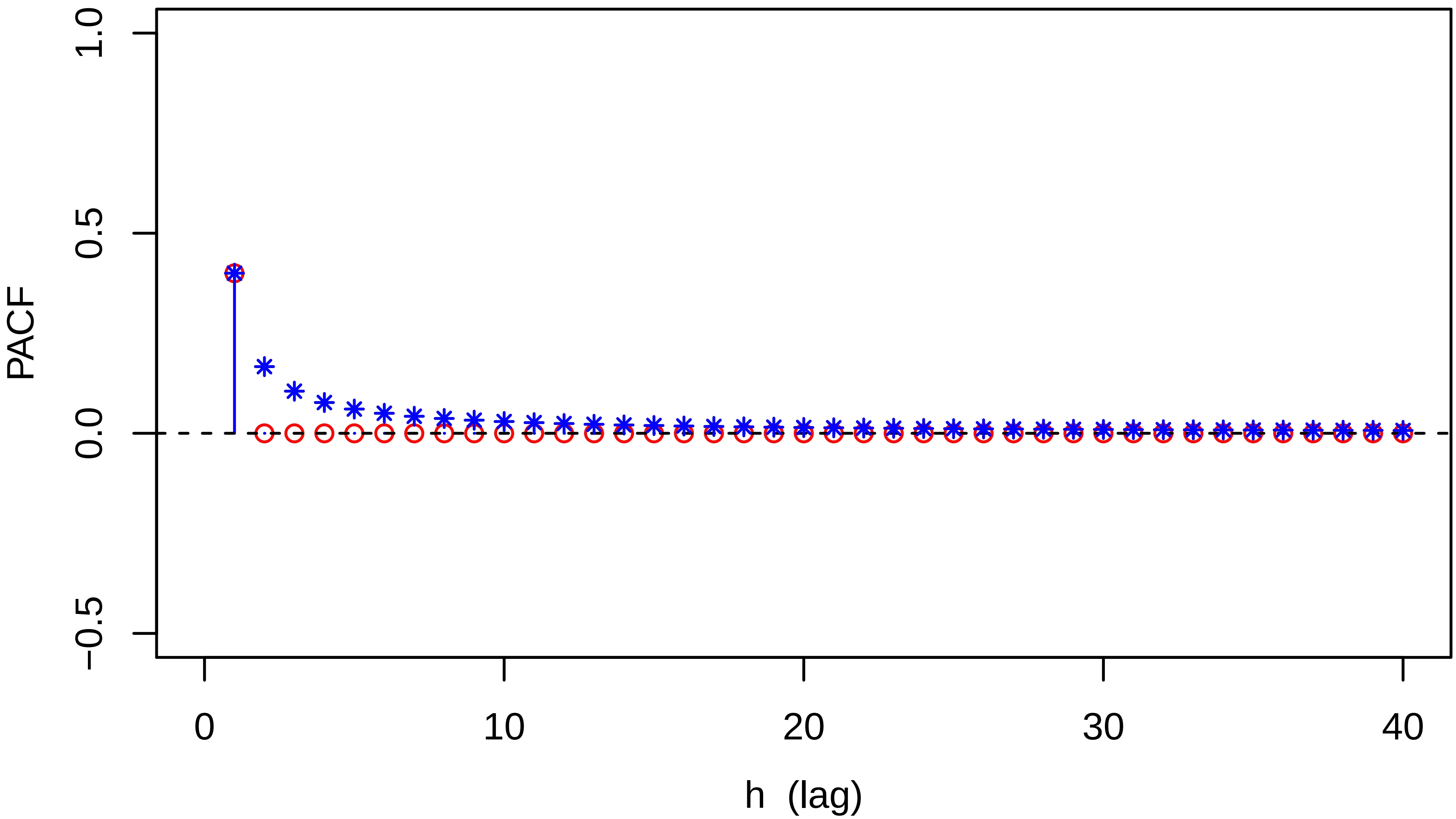



Comparison of AR (∘) and FD (∗) ACFs, $\rho(1) = 0.9$
ACF
1.0
0.5
0.0
−0.5
0
10
20
30
40
h (lag)

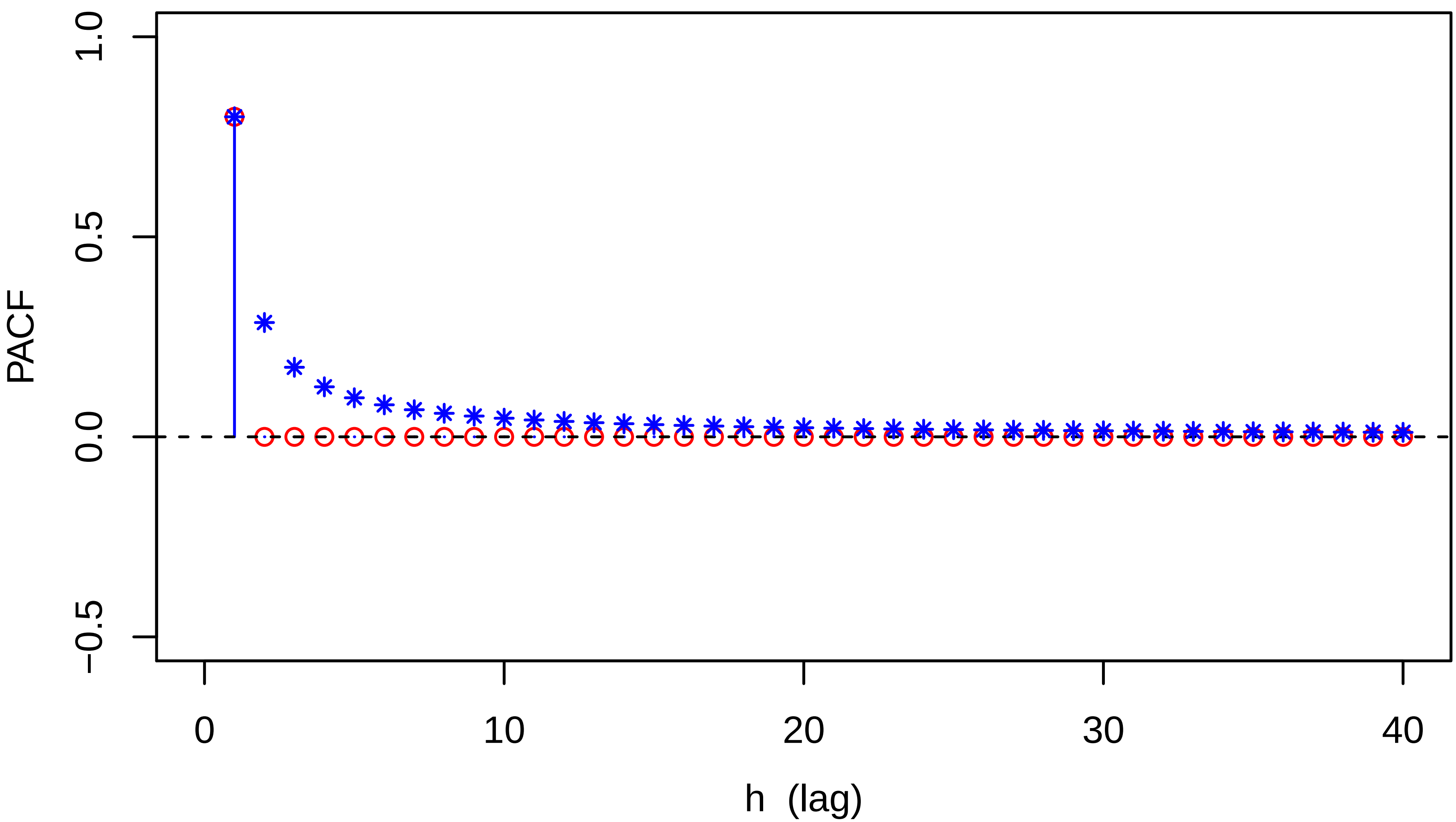
Comparison of AR (∘) and FD (∗) PACFs, $\rho(1) = 0.2$



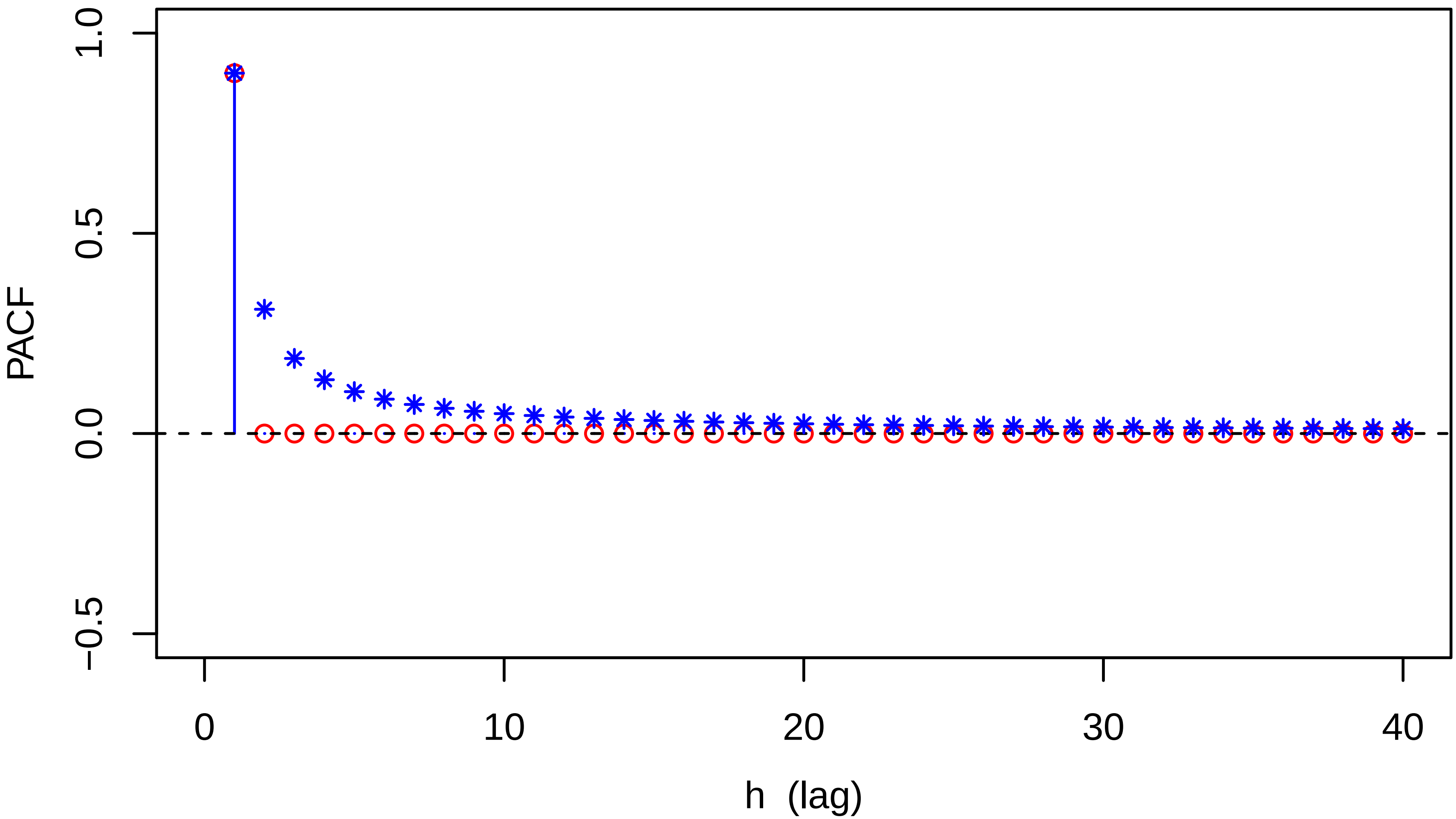
Comparison of AR ($\circ$) and FD ($*$) PACFs, $\rho(1) = 0.4$



Comparison of AR (∘) and FD (∗) PACFs, $\rho(1) = 0.8$



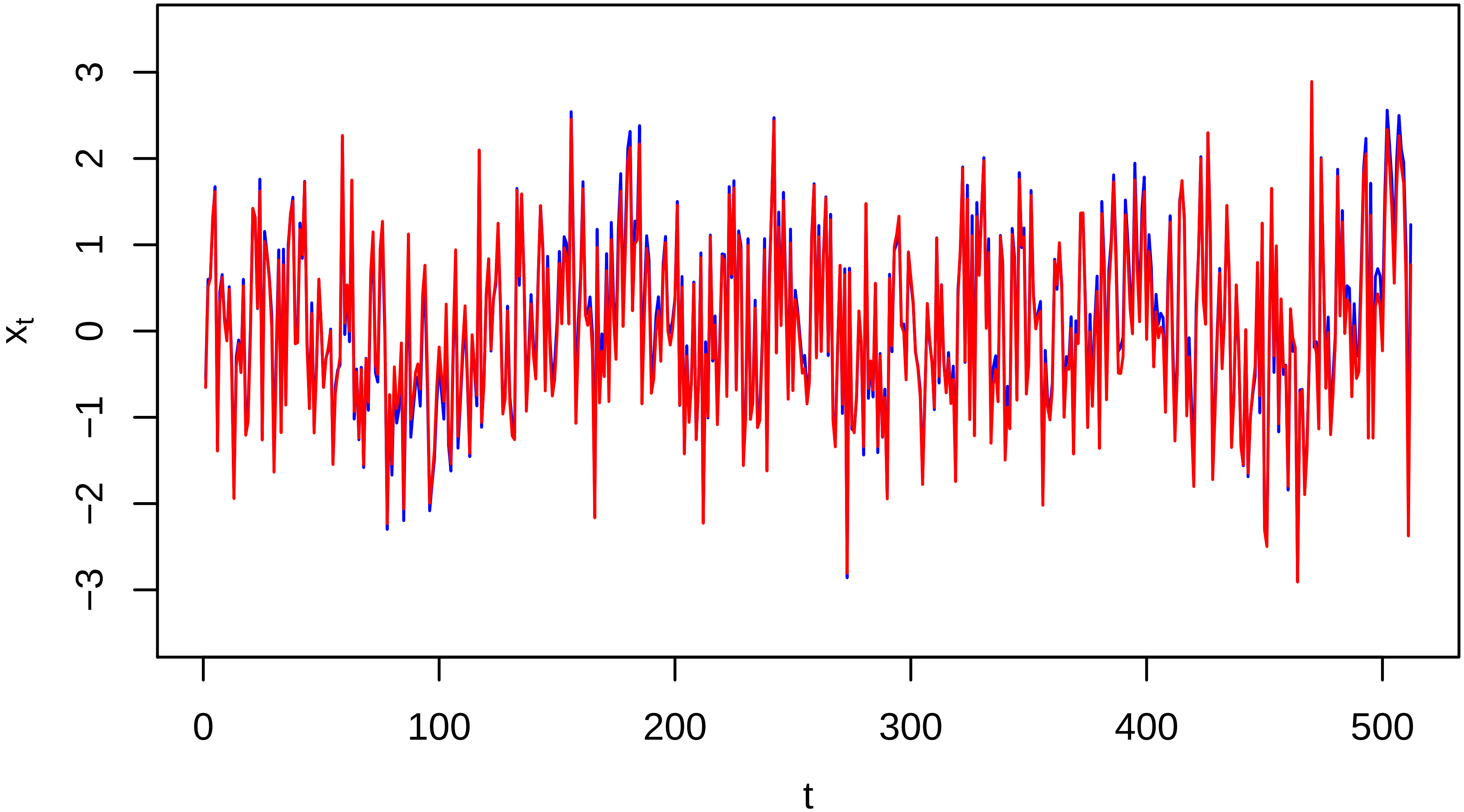
Comparison of AR (∘) and FD (∗) PACFs, $\rho(1) = 0.9$



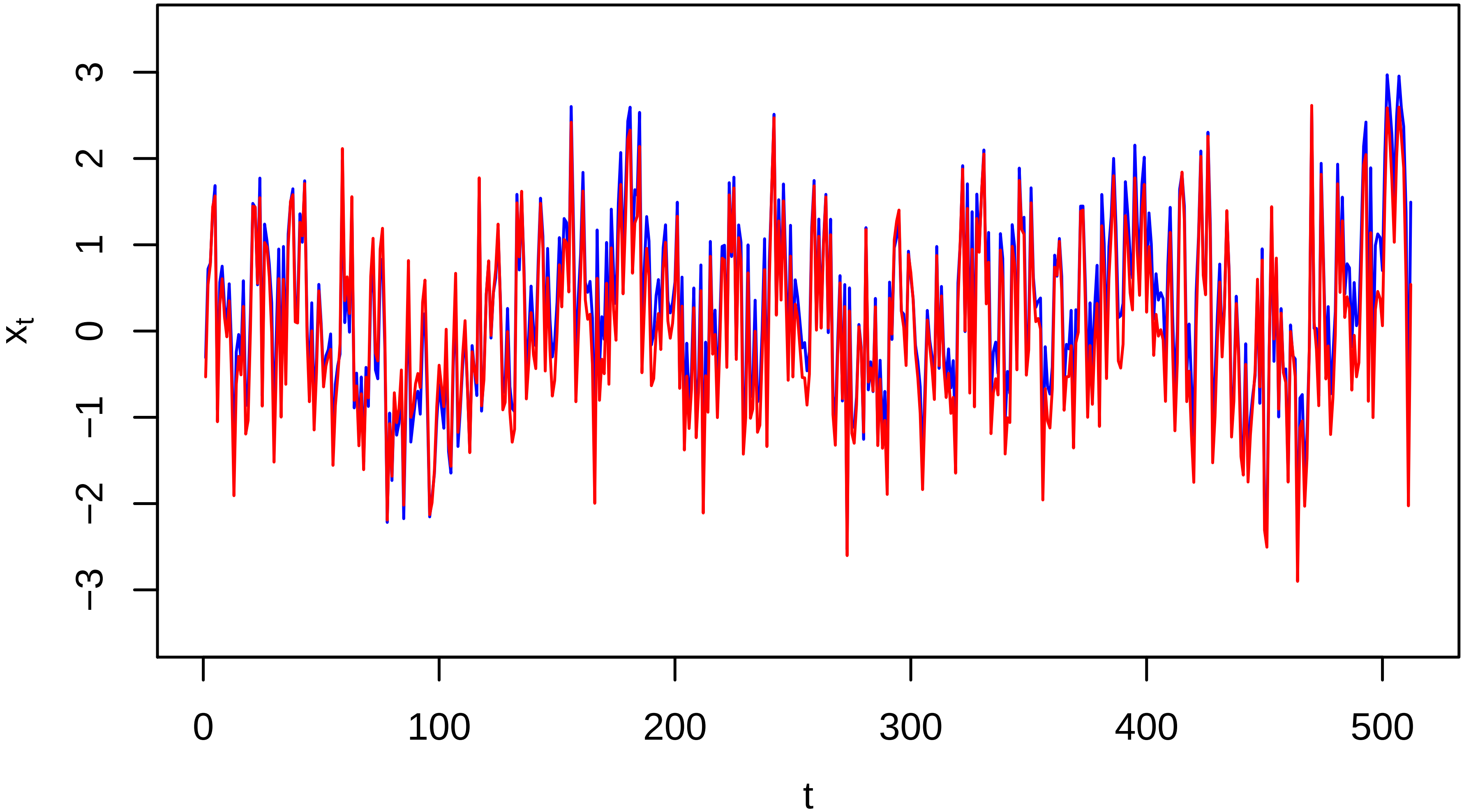
Fractionally Differenced Processes: V

- FD processes with $0 < \delta < 1/2$ are said to exhibit long-range dependence (or to have long memory)
- by contrast, terms 'short range' and 'short memory' are used to describe AR and other ARMA processes
- visual difference between short- & long-range dependence subtle
- following overheads show simulated AR and FD time series with unit variance and same $\rho(1)$ generated using circulant embedding (Davies and Harte, 1987; Gneiting, 2000; Craigmile, 2003)
- circulant embedding is an 'exact' simulation method
- series created using *same* set of Gaussian white noise deviates

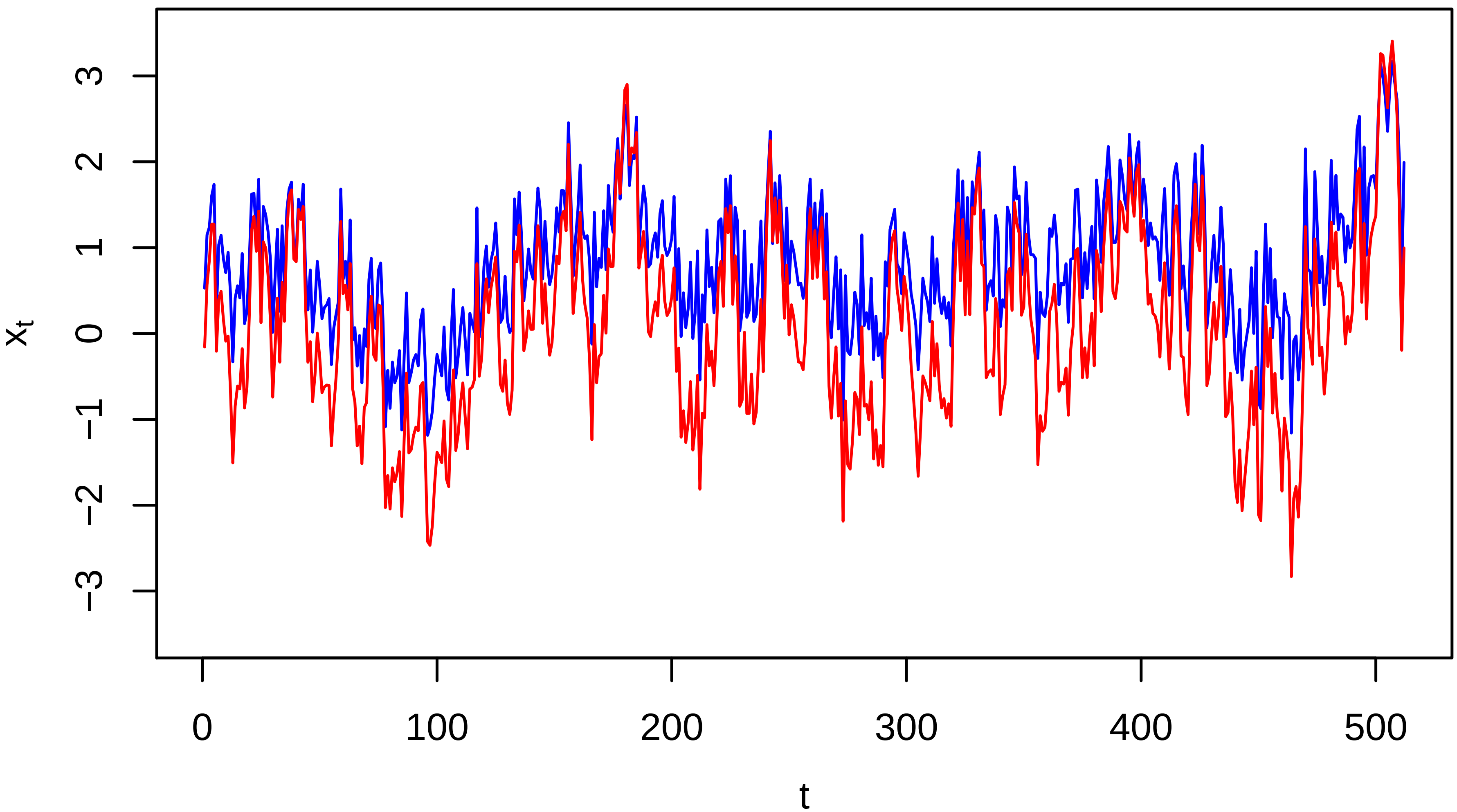
Simulated AR and FD Time Series, $\rho(1) = 0.2$



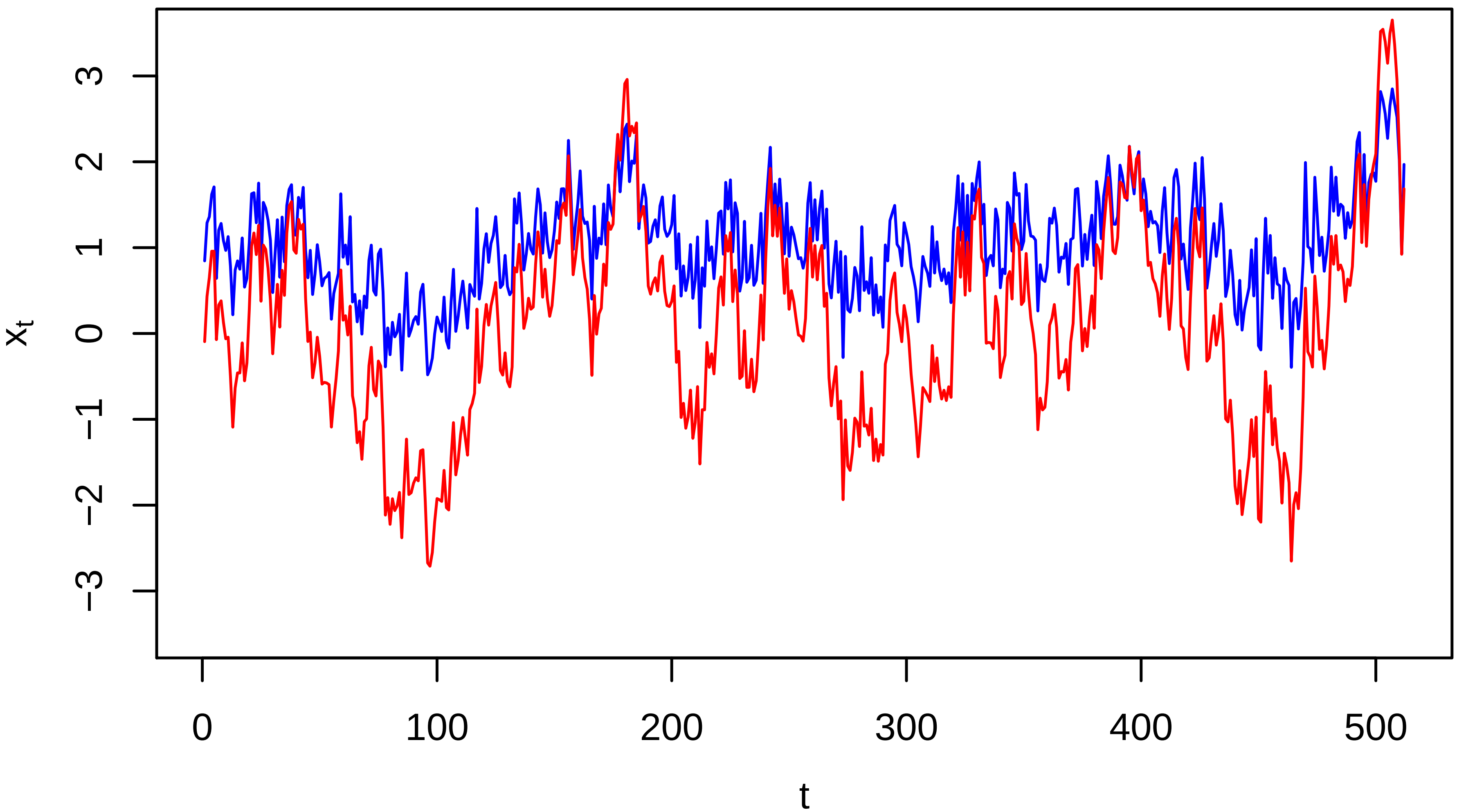
Simulated AR and FD Time Series, $\rho(1) = 0.4$



Simulated AR and FD Time Series, $\rho(1) = 0.8$



Simulated AR and FD Time Series, $\rho(1) = 0.9$



Fractionally Differenced Processes: VI

- starting from overhead XVI–3, can write
$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k Z_{t-k} = (1-B)^d Z_t$$
for $d \in (-1/2, 1/2)$ if we define
$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k B^k$$
- recalling that $\delta = -d$, we have $X_t = (1-B)^{-\delta} Z_t$
- equivalently, we can write $(1-B)^{\delta} X_t = Z_t$
- for $\delta \in [-1/2, 1/2)$, we can now define an ARIMA(p,δ,q) process via $\phi(B)(1-B)^{\delta} X_t = \theta(B) Z_t$
- note: case $\delta = -1/2$ requires special care

Fractionally Differenced Processes: VII

- ARIMA(p,δ,q) process $\phi(B)(1-B)^\delta X_t = \theta(B)Z_t$ referred to as ARFIMA(p,δ,q) process ('FI' = 'fractionally integrated')
  - sometimes called FARIMA(p,δ,q) process
- to create an ARFIMA(p,δ,q) process, take FD process $Y_t = (1-B)^{-\delta} Z_t$ and subject it to AR and MA filtering operations:

$$\theta(B)\phi^{-1}(B)Y_t = \theta(B)\phi^{-1}(B)(1-B)^{-\delta} Z_t$$

- denoting filter output $\theta(B)\phi^{-1}(B)Y_t$ by X_t, have

$$X_t = \theta(B)\phi^{-1}(B)(1-B)^{-\delta} Z_t,$$

  which, upon rearrangement, yields

$$\phi(B)(1-B)^\delta X_t = \theta(B)Z_t$$

Fractionally Differenced Processes: VIII

- given ARFIMA(p, δ, q) process Y_t with $\delta \in [-1/2, 1/2)$, can
  - form first difference to define ARFIMA(p, δ, q) process
    $$X_t = \nabla Y_t = Y_{t-1} - Y_t$$
    with $\delta \in [-3/2, -1/2)$
  - form cumulative sums to define ARFIMA(p, δ, q) process
    $$X_t = \begin{cases} X_0 + \sum_{u=1}^{t} Y_u, & t > 0, \\ X_0, & t = 0, \\ X_0 - \sum_{u=0}^{|t|-1} Y_{-u}, & t < 0 \end{cases}$$
    with $\delta \in [1/2, 3/2)$, where $X_0 = 0$ is one viable choice
- additional differences and sums define ARFIMA(p, δ, q) process for all $\delta \in \mathbb{R}$

Fractionally Differenced Processes: IX

- for $\delta \in \mathbb{R}$, take FD(δ) to be shorthand for ARFIMA(0,δ,0)
- as δ sweeps from 0 to 1, FD processes 'interpolate' between white noise ($\delta = 0$) and random walk ($\delta = 1$)
- FD process X_t with $\delta < 1/2$ is stationary
- FD process X_t with $\delta \geq 1/2$ is nonstationary, but $\nabla^d X_t$ is stationary FD process with parameter δ_s, where

$$d = \lfloor \delta + 1/2 \rfloor \text{ and } \delta_s = \delta - d$$

- FD process with $\delta = 1/2$ is an interesting special nonstationary case on the stationary/nonstationary divide
  - related to so-called $1/f$ noise (also called flicker noise)
- all ARFIMA processes are intrinsically stationary

Variance of Sample Mean Revisited: I

- suppose $X_1, \ldots, X_n$ is a portion of a stationary process with mean μ and ACVF $\{\gamma(h)\}$
- usual estimator of μ is the sample mean:

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t$$

- overhead V–5 says that

$$\text{var}\,\{\overline{X}_n\} = \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)$$

- in addition, as $n \to \infty$, if $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then

$$n\, \text{var}\,\{\overline{X}_n\} \to \sum_{h=-\infty}^{\infty} \gamma(h)$$

BD: 50, CC: 28, SS: 27

Variance of Sample Mean Revisited: II

- based upon $n \operatorname{var}\{\overline{X}_n\} \to \sum_h \gamma(h)$, have, for large n,

$$\operatorname{var}\{\overline{X}_n\} \approx \frac{v}{n}, \text{ where } v = \sum_{h=-\infty}^{\infty} \gamma(h),$$

  which is a useful approximation if $0 < v < \infty$

- Problem 10 considered three stationary processes constructed from $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, namely,
  1. $X_t = \mu + Z_1$ for which $\operatorname{var}\{\overline{X}_n\} = \sigma^2$, and approximation is useless because $v = \infty$
  2. $X_t = \mu + Z_t + Z_{t-1}$, for which $\operatorname{var}\{\overline{X}_n\} = \frac{4\sigma^2}{n} - \frac{2\sigma^2}{n^2}$ and approximation $\frac{4\sigma^2}{n}$ is useful
  3. $X_t = \mu + Z_t - Z_{t-1}$, for which $\operatorname{var}\{\overline{X}_n\} = \frac{2\sigma^2}{n^2}$, and approximation is not particularly useful because $v = 0$

Variance of Sample Mean Revisited: III

- one characterization of a stationary process with short-range dependence is that v is finite and thus, if $v \neq 0$, approximation $\text{var}\,\{\overline{X}_n\} \approx \frac{v}{n}$ is useful
- infinite v then characterizes stationary processes with long-range dependence, in which case approximation is not useful
- can show that, for an FD(δ) process with $0 < \delta < 1/2$,

$$\text{var}\,\{\overline{X}_n\} \approx \frac{c}{n^{1-2\delta}},$$

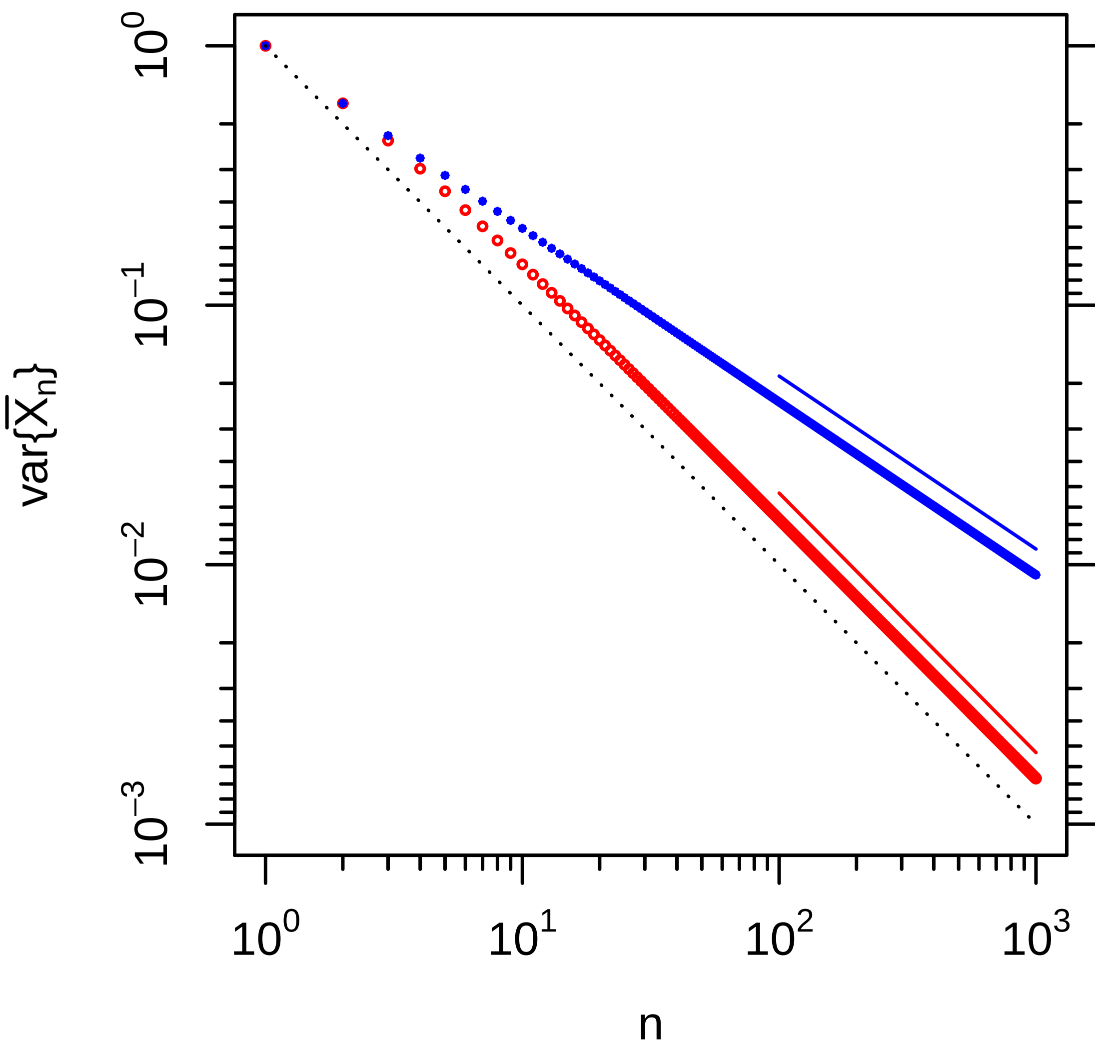
  i.e., $\text{var}\,\{\overline{X}_n\} \to 0$ as $n \to \infty$, but at a slower rate of decrease than for short-range dependence
- since $\log(\text{var}\,\{\overline{X}_n\}) \approx \log(c) - (1-2\delta)\log(n)$, plot of $y_n = \log(\text{var}\,\{\overline{X}_n\})$ versus $x_n = \log(n)$ should be approximately linear with a slope of $-1 < 2\delta - 1 < 0$ for large n

Variance of Sample Mean Revisited: IV

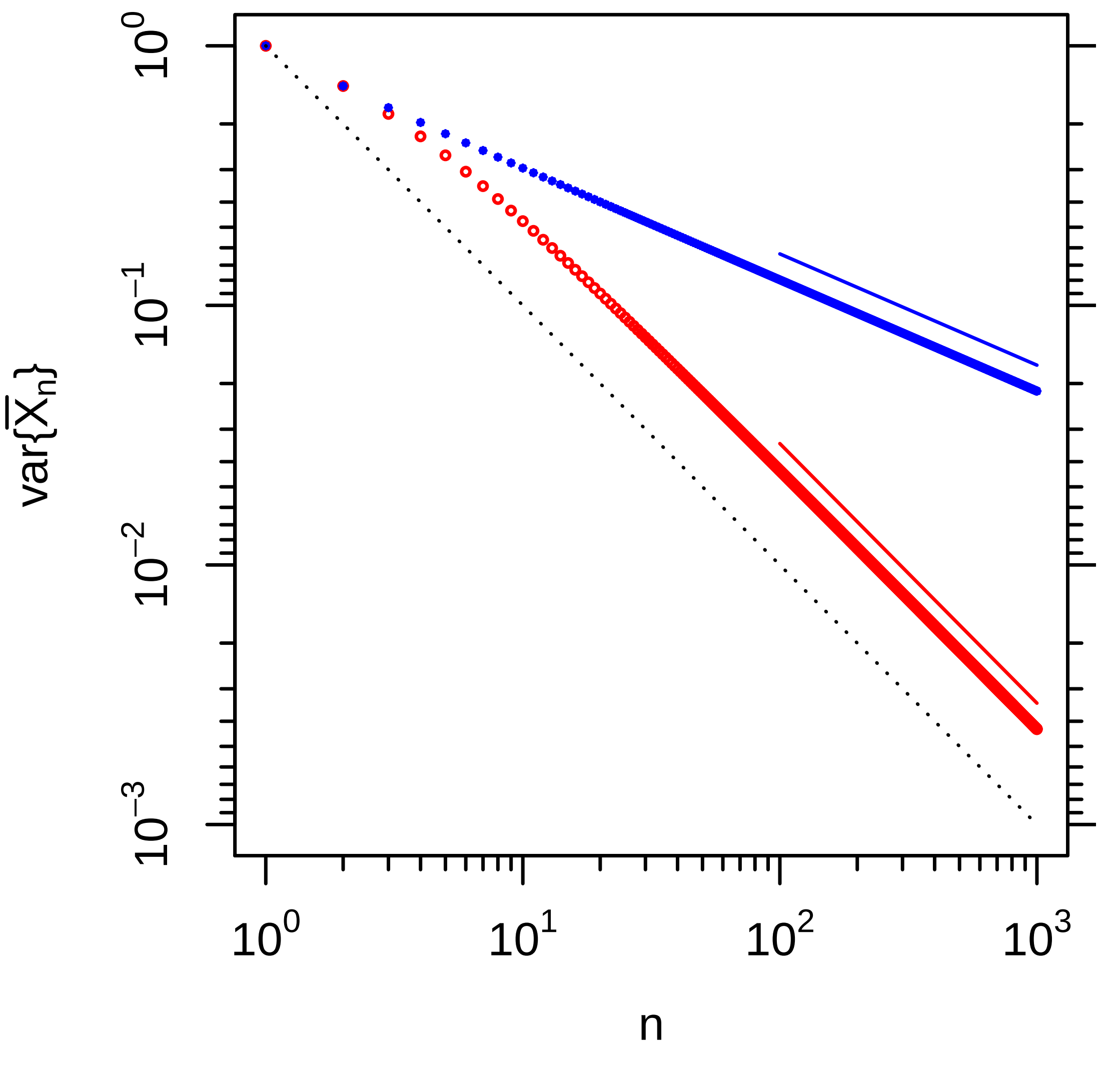
- following plots show $\text{var}\,\{\overline{X}_n\} = \frac{1}{n}\sum_{h=-(n-1)}^{n-1}\left(1 - \frac{|h|}{n}\right)\gamma(h)$ versus n on a log/log scale for AR(1) and FD processes with unit variances and with parameters ϕ and δ adjusted so that $\rho(1)$'s are identical
  - dotted black line shows $\text{var}\,\{\overline{X}_n\} = 1/n$, which is appropriate for white noise (has a slope of -1 on log/log plot)
  - thin red line above $\text{var}\,\{\overline{X}_n\}$ for AR(1) indicates large-sample rate of decay $1/n$ (has a slope of -1 on log/log plot)
  - thin blue line above $\text{var}\,\{\overline{X}_n\}$ for FD(δ) indicates large-sample rate of decay $1/n^{1-2\delta}$ (has a slope of $2\delta - 1$)
  - values of δ when $\rho(1) = \phi = 0.2$, 0.4, 0.8 and 0.9 are

$$\delta = \frac{\rho(1)}{1+\rho(1)} \doteq 0.167, \;\; 0.286, \;\; 0.444 \;\text{ and }\; 0.474$$

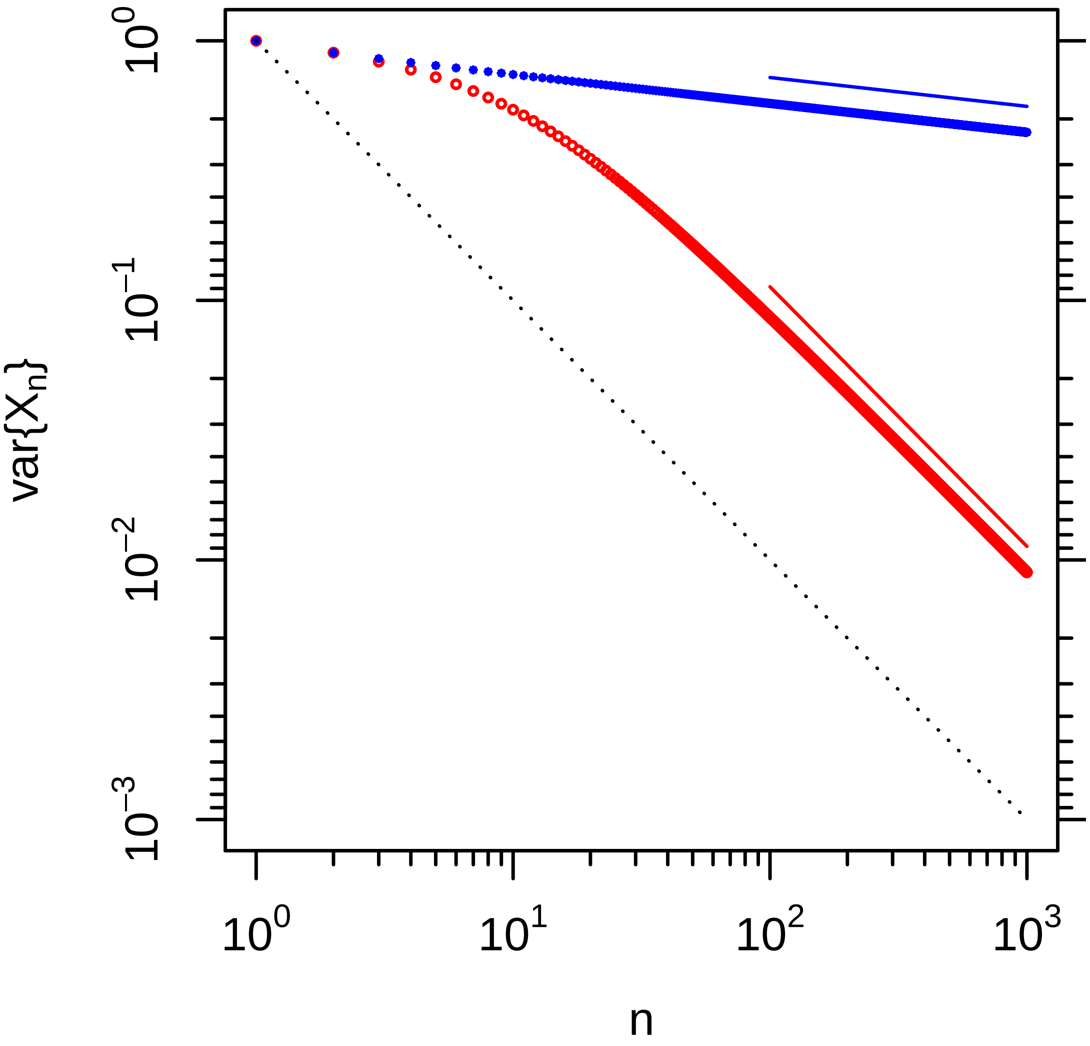
Comparison of AR (∘) and FD (∗) $\mathrm{var}\{\overline{X}_n\}$, $\rho(1) = 0.2$



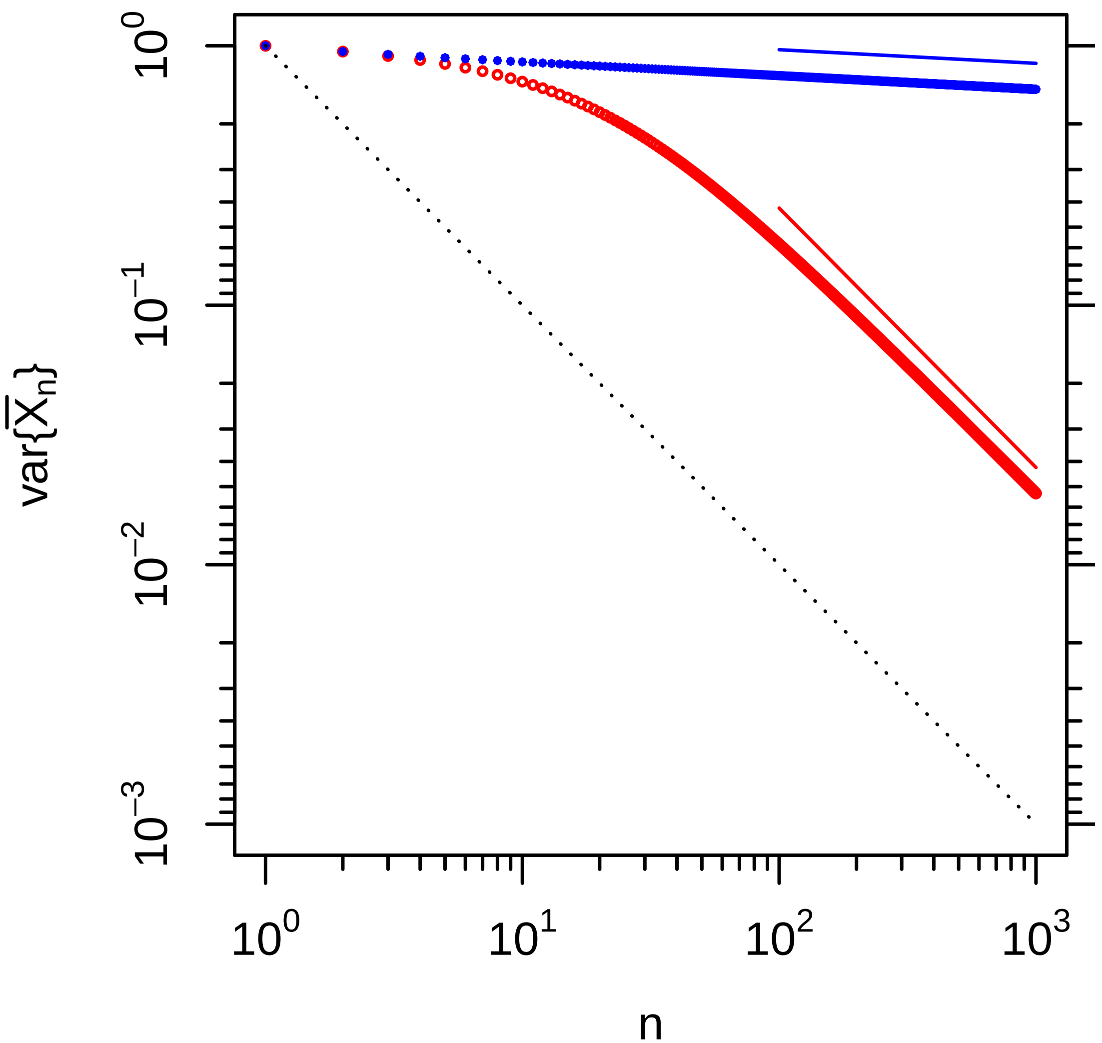
Comparison of AR (∘) and FD (∗) $\text{var}\,\{\overline{X}_n\}$**,** $\rho(1) = 0.4$



Comparison of AR (∘) and FD (∗) $\mathrm{var}\,\{\overline{X}_n\}$, $\rho(1) = 0.8$



Comparison of AR (∘) and FD (∗) $\mathrm{var}\{\overline{X}_n\}$, $\rho(1) = 0.9$



Estimation of Process Variance: I

- can be difficult to estimate variance σ_X^2 for process X_t with long-range dependence
- to understand why, assume first $\mu = E\{X_t\}$ is known
- can estimate σ_X^2 using

$$\tilde{\sigma}_X^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)^2$$

- estimator above is unbiased: $E\{\tilde{\sigma}_X^2\} = \sigma_X^2$
- if μ is unknown (more common case), can estimate σ_X^2 using

$$\hat{\sigma}_X^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X}_n)^2$$

Estimation of Process Variance: II

- can show that $E\{\hat{\sigma}^2_X\} = \sigma^2_X - \text{var}\,\{\overline{X}_n\}$ (do it!)
- implies $0 \le E\{\hat{\sigma}^2_X\} \le \sigma^2_X$ because $\text{var}\,\{\overline{X}_n\} \ge 0$
  - in what surely is the shortest article in a well-known statistical journal, David (1985) shows above true for $\hat{\sigma}^2_X$ from *any* collection of dependent RVs with same mean and variance
- since $\text{var}\,\{\overline{X}_n\} \approx \frac{c}{n^{1-2\delta}}$ for FD(δ) process with $0 < \delta < 1/2$, get that $E\{\hat{\sigma}^2_X\} \to \sigma^2_X$ as $n \to \infty$ ... but, for any

  $\epsilon > 0$ (say, $0.00\cdots01$) and sample size n (say, $n = 10^{10^{10}}$),

  there is some FD process $\{X_t\}$ with δ close to $1/2$ such that

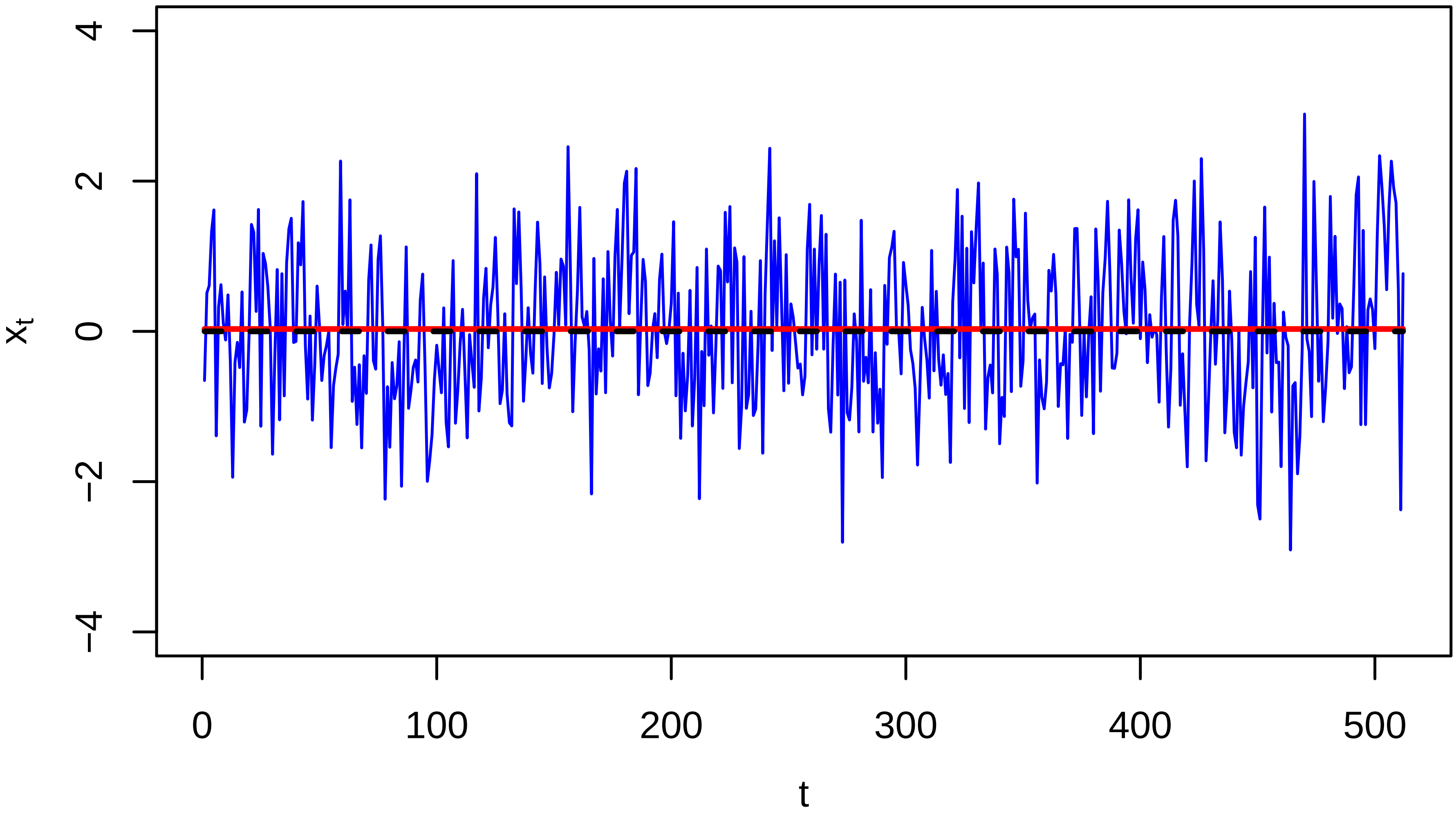
  $$E\{\hat{\sigma}^2_X\} < \epsilon \cdot \sigma^2_X;$$

  i.e., in general, $\hat{\sigma}^2_X$ can be *badly* biased even for *very* large n

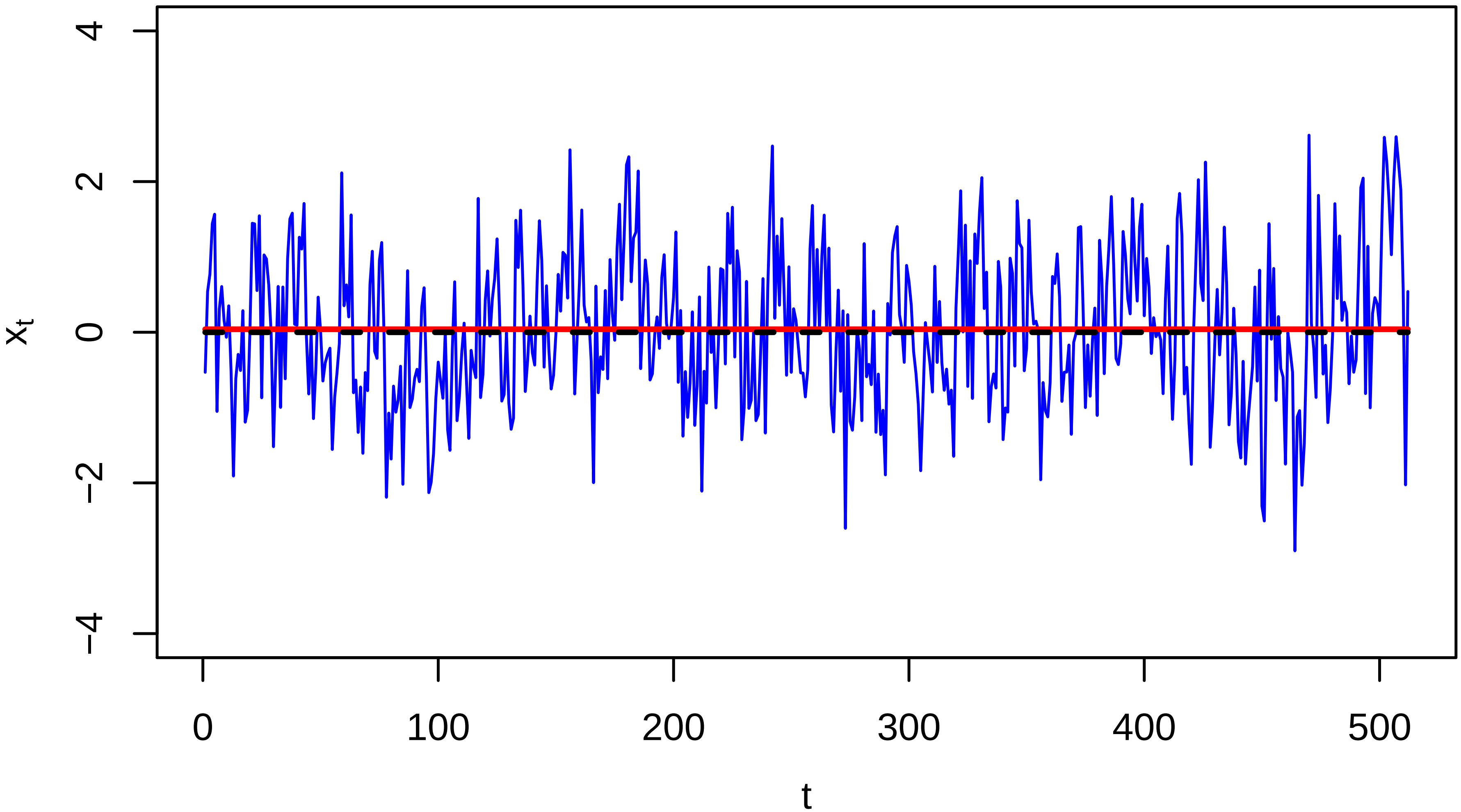
Estimation of Process Variance: III

- reconsider four previously shown simulated AR(1) and FD time series, each with sample size $n = 512$ drawn from processes with mean $\mu = 0$ and variance $\sigma_X^2 = 1$
- blue horizontal lines show process means $\mu = 0$
- red horizontal lines show sample means $\overline{X}_n$

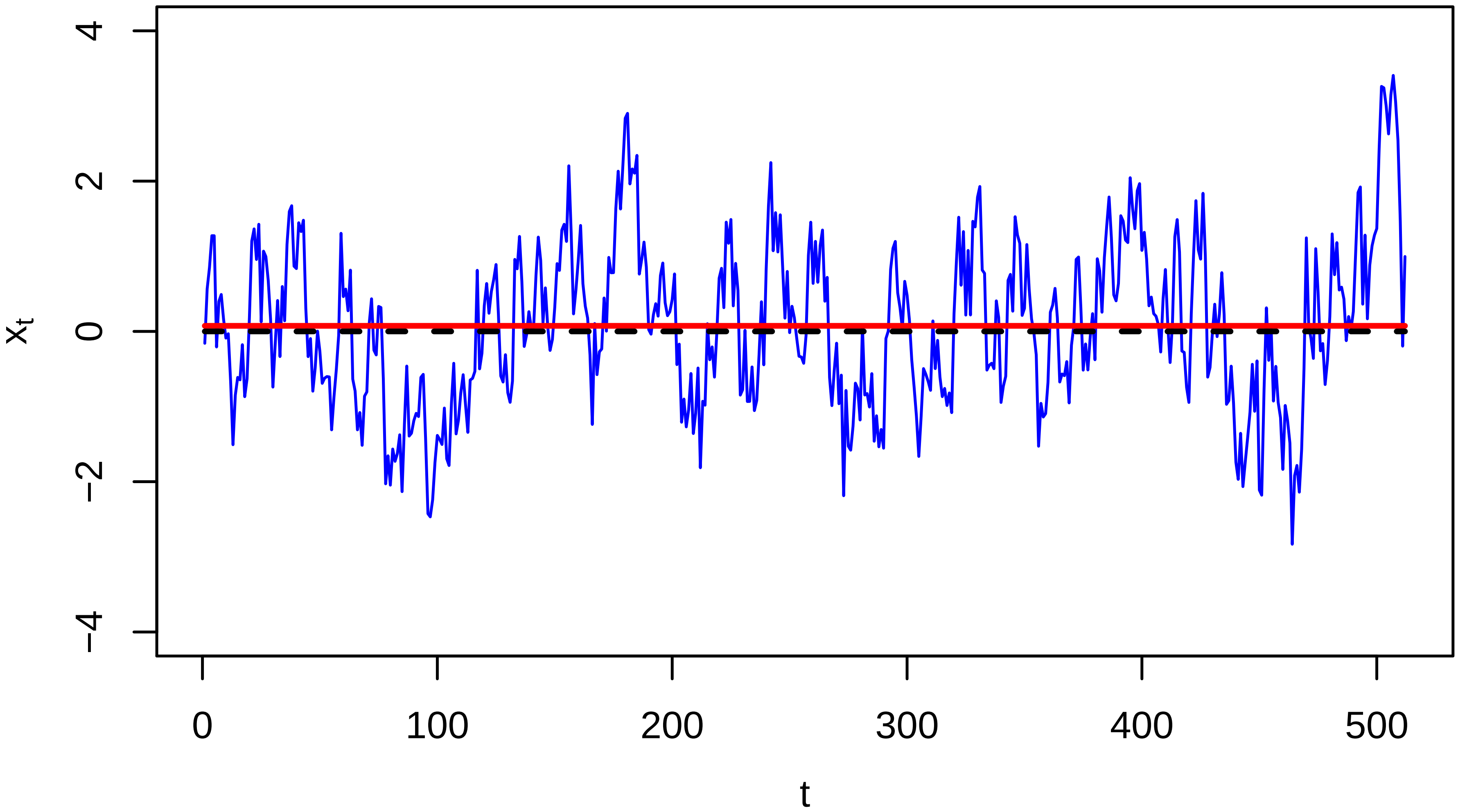
Simulated AR Time Series, $\phi = 0.2$



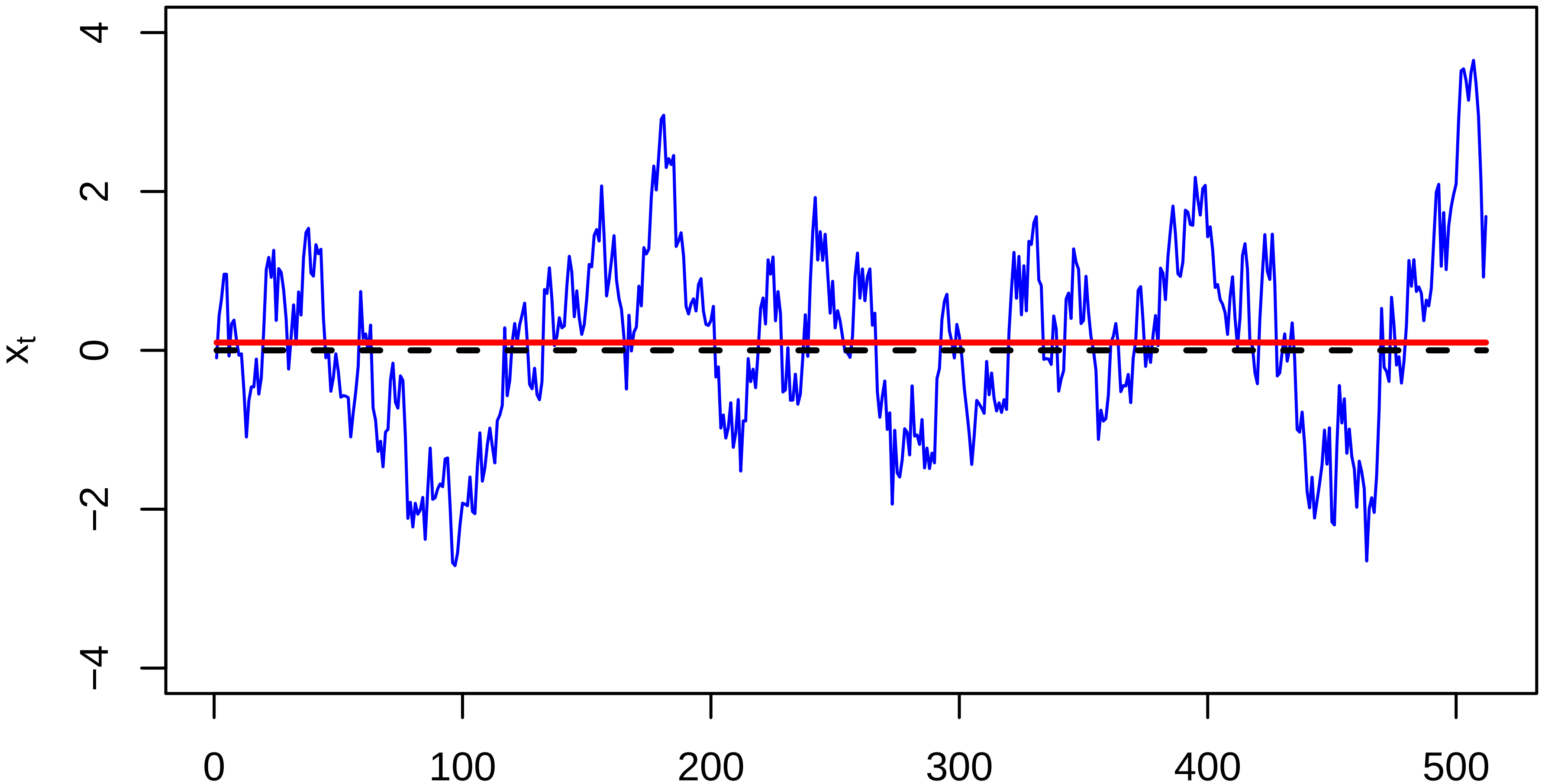
Simulated AR Time Series, $\phi = 0.4$



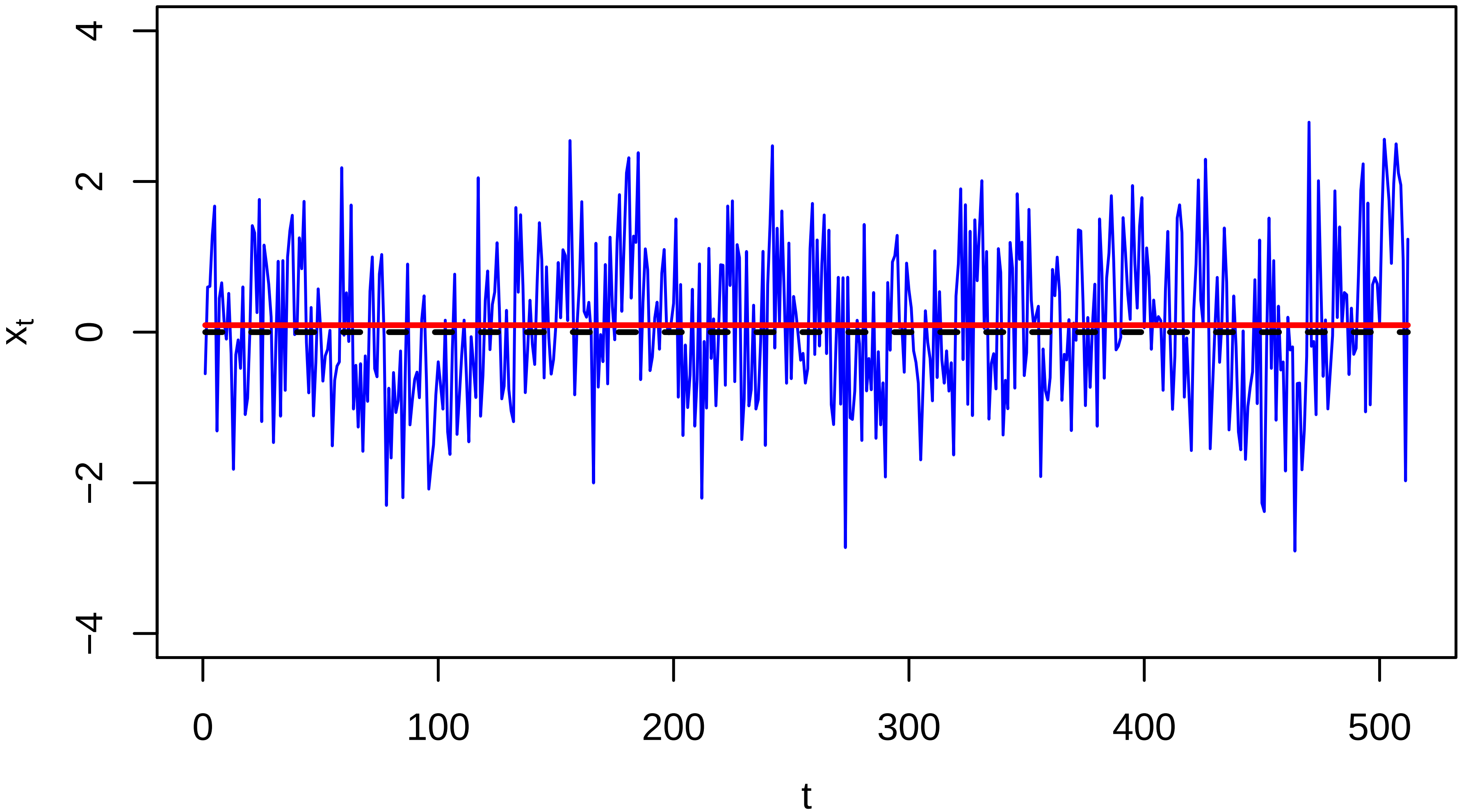
Simulated AR Time Series, $\phi = 0.8$



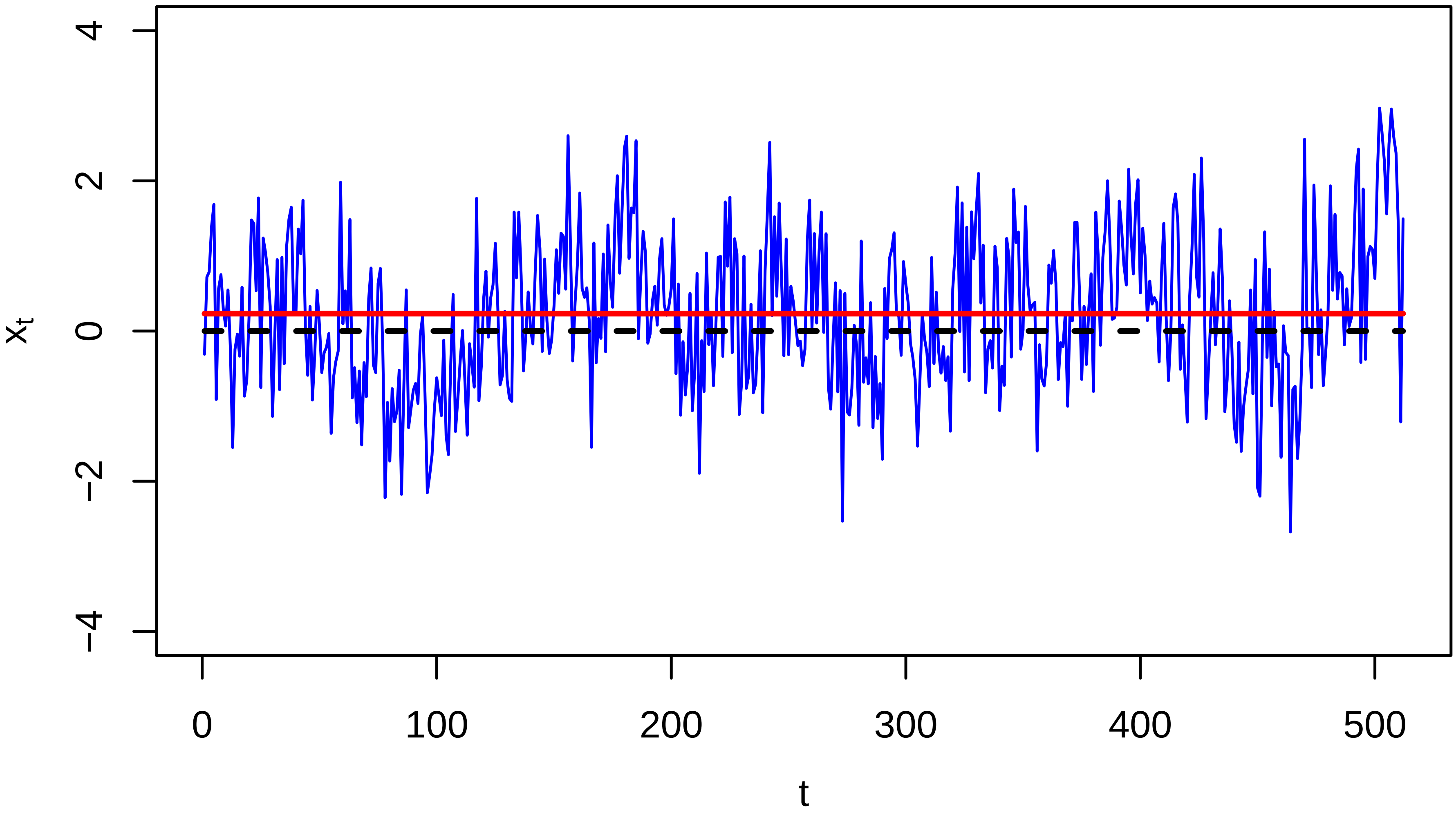
Simulated AR Time Series, $\phi = 0.9$



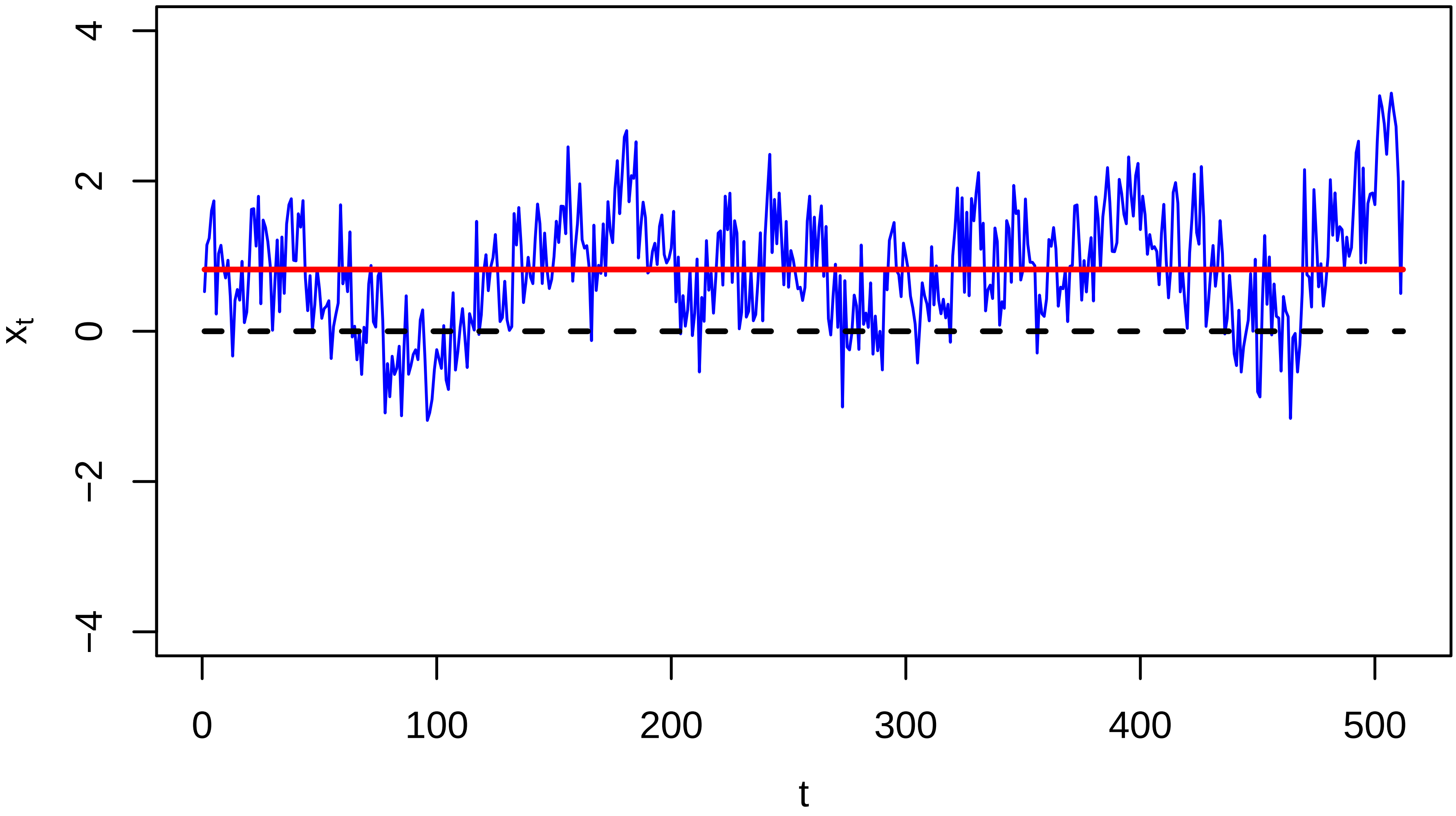
Simulated FD Time Series, $\delta \doteq 0.167$



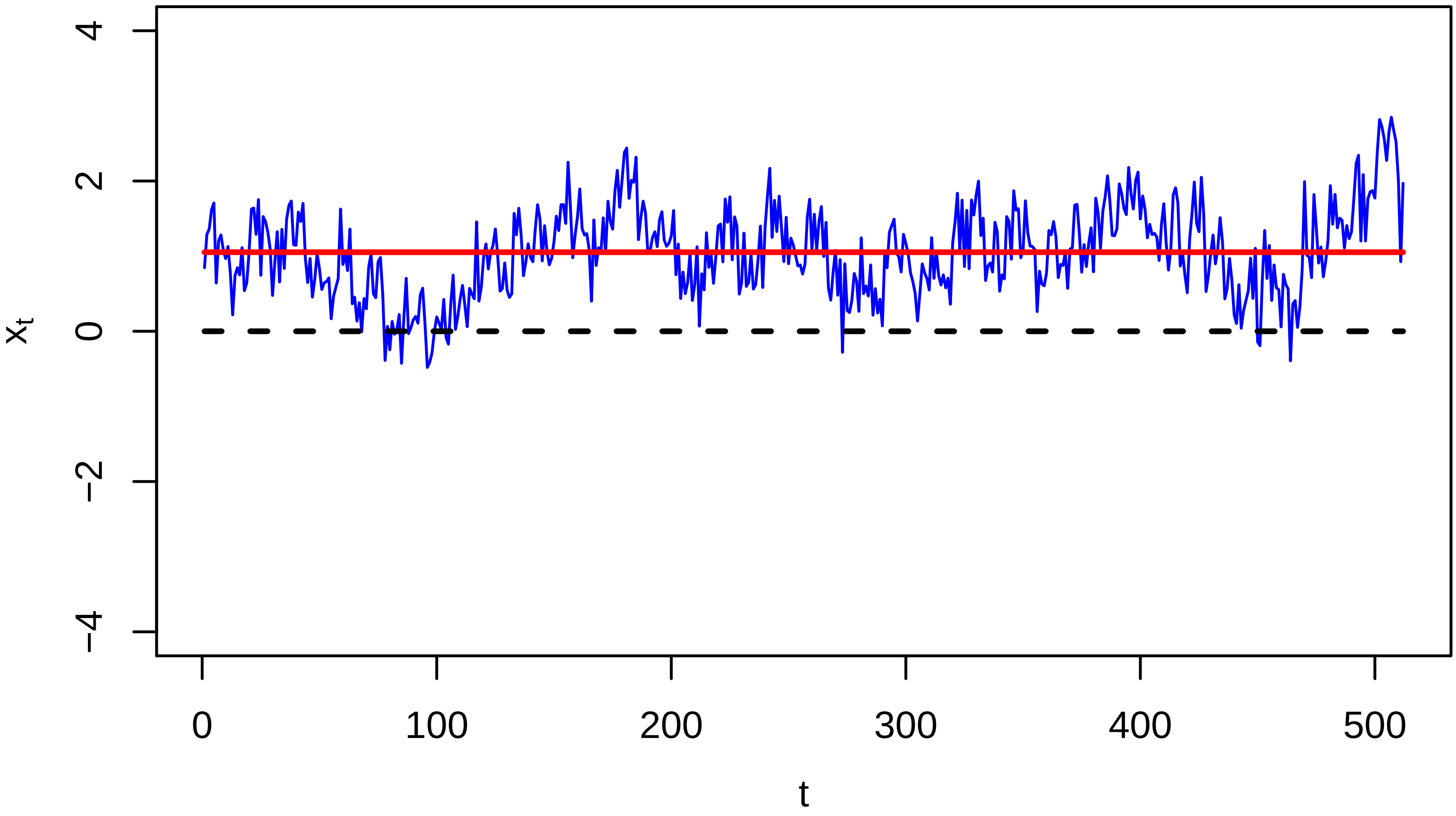
Simulated FD Time Series, $\delta \doteq 0.286$



Simulated FD Time Series, $\delta \doteq 0.444$



Simulated FD Time Series, $\delta \doteq 0.474$



Estimation of Process Variance: IV

process	σ_X^2	$\tilde{\sigma}_X^2$	$\hat{\sigma}_X^2$	$E\{\hat{\sigma}_X^2\}$	$n_{1\%}$
AR(1), $\phi = 0.2$	1.00	1.04	1.03	1.00	150
AR(1), $\phi = 0.4$	1.00	1.04	1.03	1.00	233
AR(1), $\phi = 0.8$	1.00	1.17	1.16	0.98	896
AR(1), $\phi = 0.9$	1.00	1.35	1.34	0.96	1891
FD(0.167)	1.00	1.07	1.06	0.99	872
FD(0.286)	1.00	1.10	1.04	0.94	36532
FD(0.444)	1.00	1.29	0.61	0.52	$\approx 7 \times 10^{17}$
FD(0.474)	1.00	1.46	0.35	0.30	$\approx 6 \times 10^{37}$

- in above, $n_{1\%}$ is the smallest sample size such that

$$\sigma_X^2 - E\{\hat{\sigma}_X^2\} \leq 0.01,$$

i.e., bias in sample variance no more than 1% of true variance

Estimation of Process Variance: V

- under assumption that X_t is well-modeled by an FD(δ) process, Beran (1994) outlines a procedure to de-bias $\hat{\sigma}^2_X$ based upon

$$E\{\hat{\sigma}^2_X\} = \sigma^2_X - \mathrm{var}\,\{\overline{X}_n\} = \sigma^2_X(1 - \mathrm{var}\,\{\overline{X}_n/\sigma_X\})$$

- first step is to get an estimate, say $\hat{\delta}$, of δ (MLE is one option)
- given $\hat{\delta}$, form correction factor $C_{\hat{\delta}} = 1 - \widehat{\mathrm{var}}\,\{\overline{X}_n/\sigma_X\}$, where

$$\widehat{\mathrm{var}}\,\{\overline{X}_n/\sigma_X\} = \frac{1}{n}\sum_{h=-(n-1)}^{n-1}\left(1 - \frac{|h|}{n}\right)\hat{\rho}(h),$$

and $\hat{\rho}(h)$ is parametric estimator of ACF based upon $\hat{\delta}$:

$$\hat{\rho}(h) = \hat{\rho}(h-1)\frac{h + \hat{\delta} - 1}{h - \hat{\delta}},\ \ h = 1, 2, \ldots,\ \ \text{with}\ \ \hat{\rho}(0) \stackrel{\text{def}}{=} 1$$

Estimation of Process Variance: VI

- approximately unbiased estimator of σ_X^2 given by $\hat{\sigma}_X^2/C_{\hat{\delta}}$
- as alternative to this semiparametric estimator, consider parametric estimator based on MLEs of δ and σ^2, where latter is variance of $Z_t \sim \mathrm{WN}(0, \sigma^2)$ used to construct FD process:

$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(k+\delta)}{\Gamma(k+1)\Gamma(\delta)} Z_{t-k}$$

- letting $\hat{\sigma}^2$ be MLE for σ^2, can use relationship

$$\sigma_X^2 = \sigma^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)}$$

  to obtain estimator for σ_X^2 by plugging in $\hat{\sigma}^2$ & $\hat{\delta}$ for σ^2 & δ
- relative merits of semiparametric and parametric estimators seem to be an open question

Maximum Likelihood Estimation: I

- $-2\times$ log of likelihood for Gaussian zero-mean stationary time series $\boldsymbol{X}_n = [X_n, \ldots, X_1]'$ with covariance matrix Γ_n given by
$$-2\ln\left(L(\Gamma_n)\right) = n\ln\left(2\pi\right) + \ln\left(\det\Gamma_n\right) + \boldsymbol{X}_n'\Gamma_n^{-1}\boldsymbol{X}_n$$
(overheads XIII–22 & XIII–116)
- for ARFIMA(p, δ, q) process, parameters $\boldsymbol{\phi}$, δ, $\boldsymbol{\theta}$ & 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 $-2\ln\left(L(\Gamma_n)\right)$ is minimized
- as special case, consider MLEs for δ and σ^2 for FD(δ) process represented by
$$X_t = \sum_{k=0}^{\infty}\frac{\Gamma(k+\delta)}{\Gamma(k+1)\Gamma(\delta)}Z_{t-k}, \text{ where } Z_t \sim \mathrm{WN}(0, \sigma^2)$$

Maximum Likelihood Estimation: II

- using same argument as before (overheads XIII–116 to 122), arrive at profile likelihood for δ:

$$-2\ln\big(L(\delta)\big) = n + n\ln(2\pi/n) + n\ln(S(\delta)) + \sum_{j=1}^{n} \ln(r_{j-1}),$$

where

$$S(\delta) = \sum_{j=1}^{n} \frac{(X_j - \widehat{X}_j)^2}{r_{j-1}}$$

and r_{j-1} is mean square error of predictor $\widehat{X}_j$ when σ^2 is set artificially to unity (actually just need $r_{j-1} \propto \text{MSE}\{\widehat{X}_j\}$)

- scheme is to find minimizer $\hat{\delta}$ of $-2\ln\big(L(\delta)\big)$, i.e., the MLE for δ, and then to get MLE for σ^2, namely, $\hat{\sigma}^2 = S(\hat{\delta})/n$

Maximum Likelihood Estimation: III

- to evaluate $-2\ln\big(L(\delta)\big)$, here are the steps we need to take

  1. set
     $$r_0 = \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)}, \quad \boldsymbol{\phi}_1 = \phi_{1,1} = \frac{\delta}{1-\delta} \text{ and } r_1 = r_0(1-\phi_{1,1}^2)$$
     note: r_0 is variance of FD(δ) process with σ^2 set to unity
  2. use modified version of L–D recursions to get coefficients $\boldsymbol{\phi}_{j-1}$ for best linear predictor $\widehat{X}_j$ of X_j given $X_{j-1}, \ldots, X_1$ along with normalized mean square error r_{j-1}; in particular, for $j = 2, \ldots, n-1$ with $\boldsymbol{\phi}_{j-1}$ and r_{j-1} given,

  2a. set $\phi_{j,j} = \frac{\delta}{j-\delta}$ (note: no need to use ACVF!)

  2b. set rest of $\boldsymbol{\phi}_j$ using $\boldsymbol{\phi}_{j-1}$ and $\phi_{j,j}$ (usual 2nd step of L–D)

  2c. set $r_j = r_{j-1}(1-\phi_{j,j}^2)$ (usual 3rd step of L–D)

Maximum Likelihood Estimation: IV

3. compute 1-step-ahead predictions

$$\widehat{X}_{j+1} = \sum_{k=1}^{j} \phi_{j,k} X_{j-k+1}, \quad 1 \le j < n-1$$

along with innovations $U_{j+1} = X_{j+1} - \widehat{X}_{j+1}$ (recall that $\widehat{X}_1 \stackrel{\text{def}}{=} 0$ and $U_1 = X_1$)

- now have all the pieces needed to compute

$$-2\ln\big(L(\delta)\big) = n + n\ln(2\pi/n) + n\ln(S(\delta)) + \sum_{j=1}^{n} \ln(r_{j-1}),$$

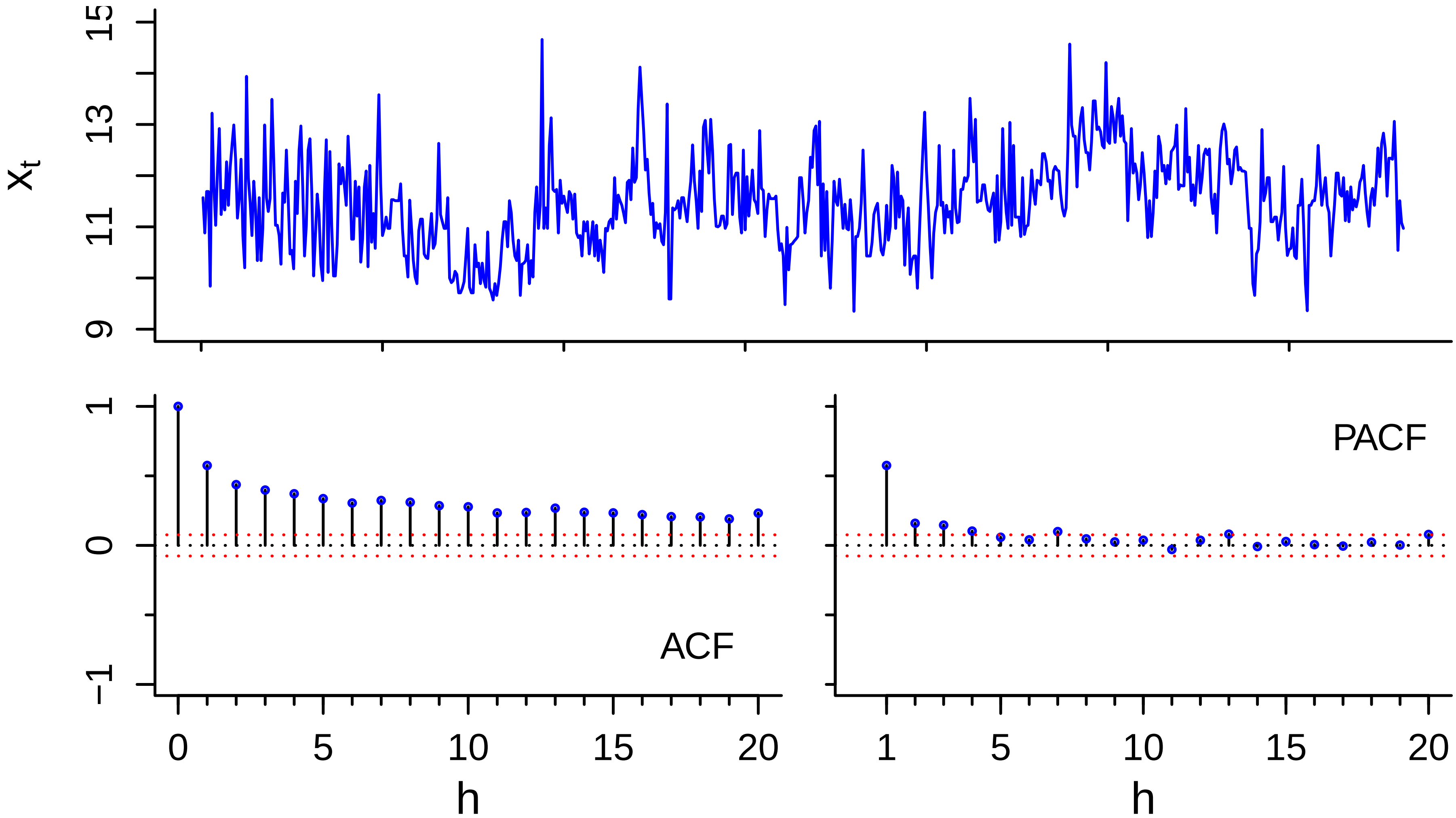
and

$$S(\delta) = \sum_{j=1}^{n} \frac{(X_j - \widehat{X}_j)^2}{r_{j-1}}$$

Maximum Likelihood Estimation: V

- given ability to compute $-2\ln\big(L(\delta)\big)$, can use nonlinear optimization scheme to find minimizer $\hat{\delta}$ (the desired MLE), from which we can then get MLE $\hat{\sigma}^2$
- can use usual theory for MLEs to get confidence intervals for unknown δ and σ^2 if so desired
- practical problem with scheme is its computational complexity: becomes unfeasible as n gets large ($n = 1000$ can be painful; more pain when dealing with ARFIMA rather than just FD)
- minor industry devoted to getting good approximations to MLEs that work well for large n
- `R` function `fracdiff` in package `fracdiff` uses fast, but accurate, approximation due to Haslett and Raftery (1989)

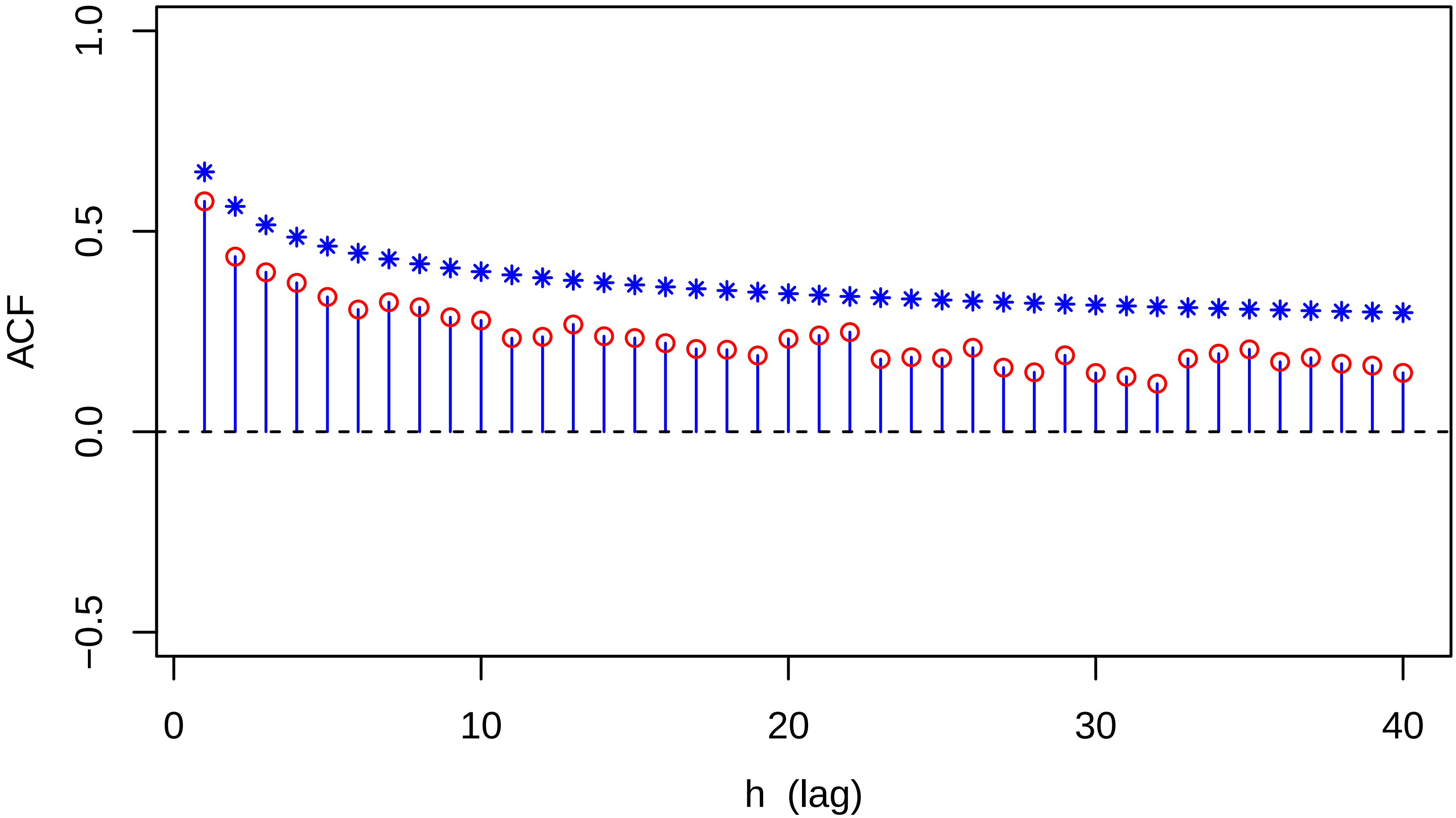
Nile River Minima Series, 622 to 1284



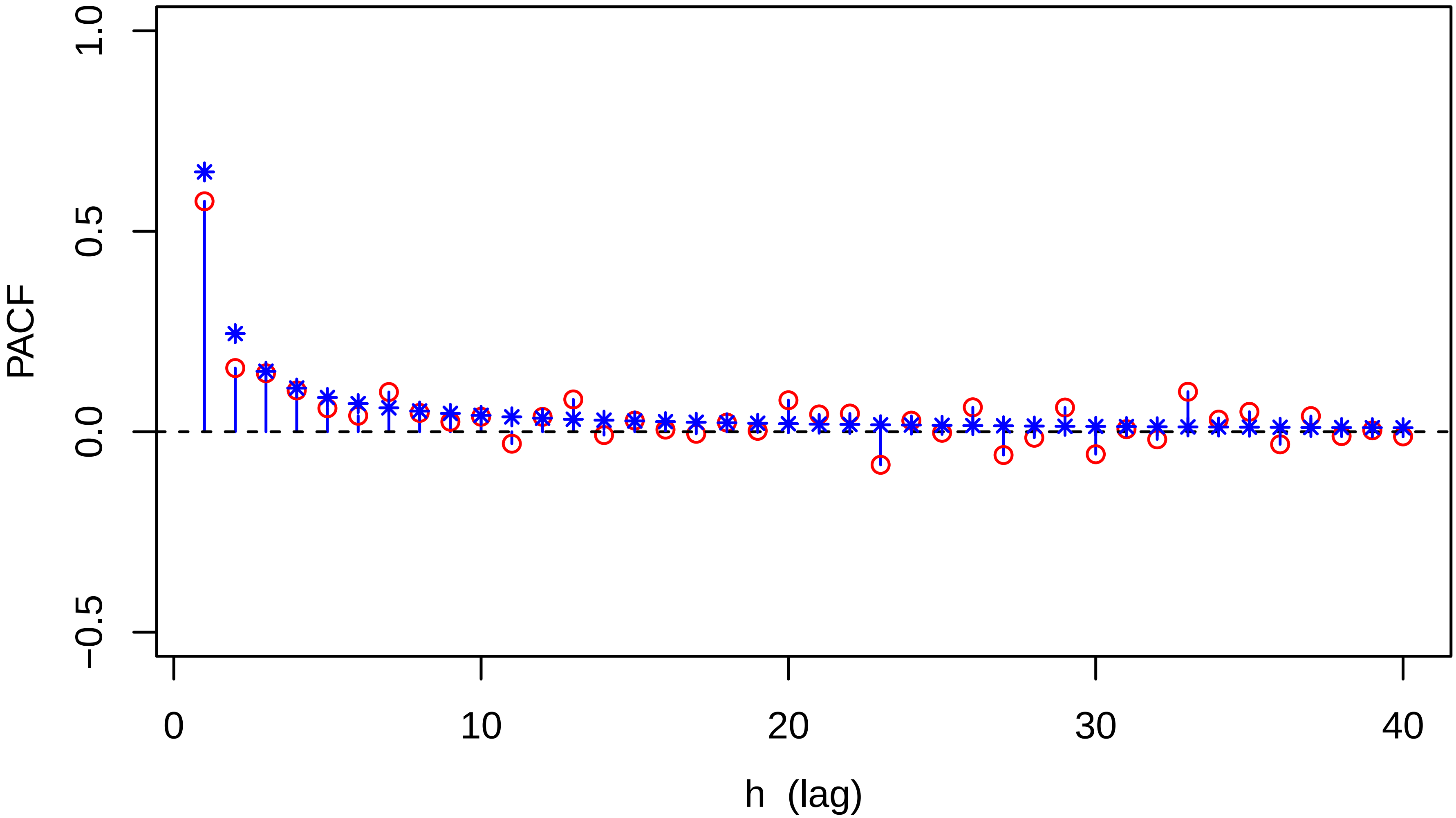
Nile River Minima as FD Process: I

- consider series of yearly minimum levels of Nile River measured at Roda gauge near Cairo from 622 to 1284
- ACF and PACF suggest that FD model might be appropriate (in fact, series regarded as classic example exhibiting long-range dependence)
- using `fracdiff` to fit FD(δ) yields $\hat{\delta} \doteq 0.393$
- let's see how well ACF and PACF for FD(0.393) process match up with sample ACF and PACF

Sample & FD ACF for Nile River



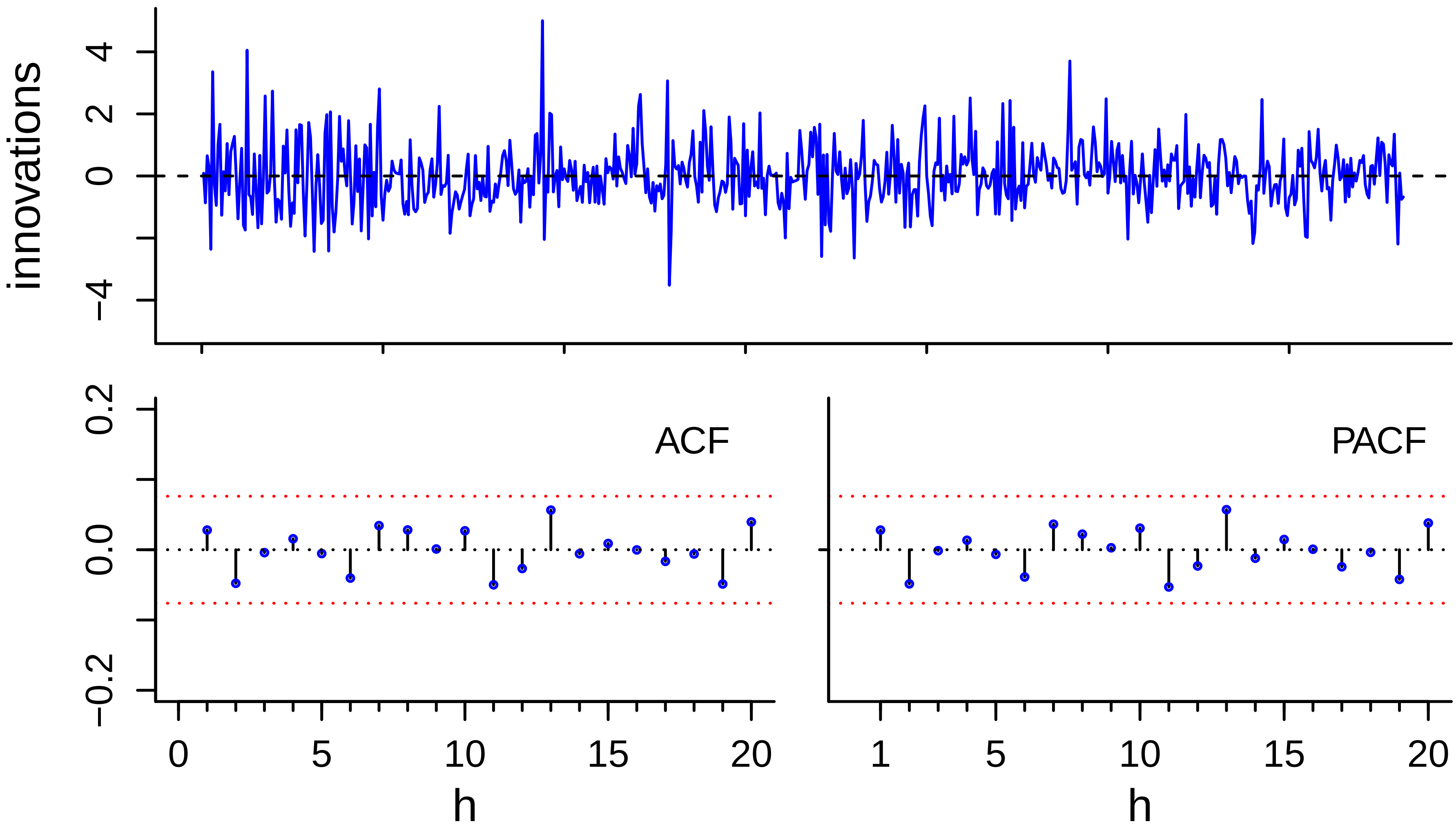
Sample & FD PACF for Nile River



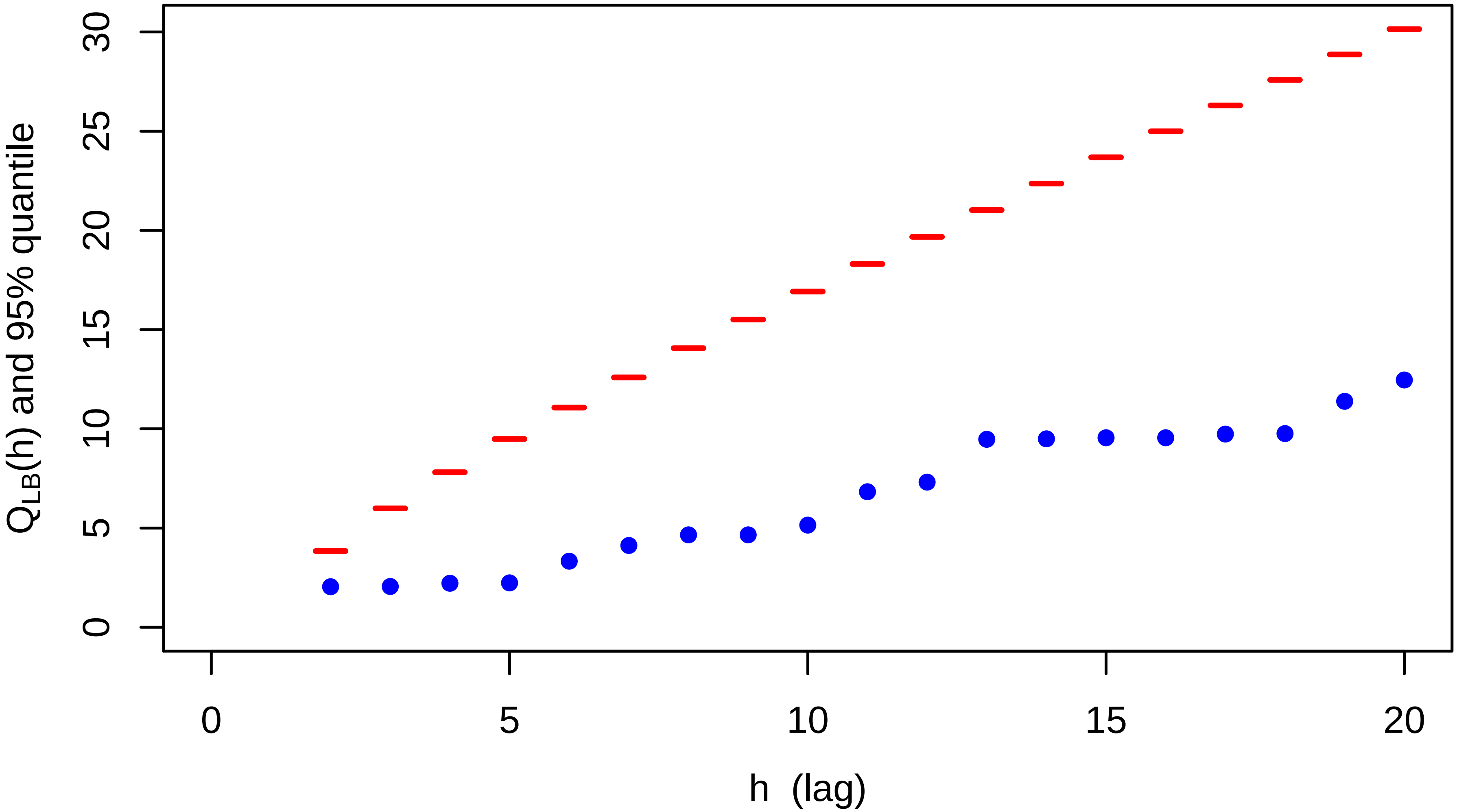
Nile River Minima as FD Process: II

- key component of profile likelihood is sum of squares of normalized innovations:
$$S(\delta) = \sum_{j=1}^{n} \frac{U_j^2}{r_{j-1}} \text{ where } U_j = X_j - \widehat{X}_j,$$
where $\widehat{X}_j$ is best linear predictor based on $X_{j-1}, \ldots, X_1$
- under correct model, innovations U_j are uncorrelated, but with variances that depend on j and that are proportional to r_{j-1}
- thus $\widehat{W}_j \stackrel{\text{def}}{=} U_j/\sqrt{r_{j-1}}$ should resemble a white noise process if our model is correct (i.e., uncorrelated with constant variance)
- following plots and tests look at normalized innovations $\widehat{W}_j$ associated with fitted FD model for Nile River series

Normalized Innovations $\widehat{W}_j$ for Nile River



Portmanteau Tests for $\widehat{W}_j$

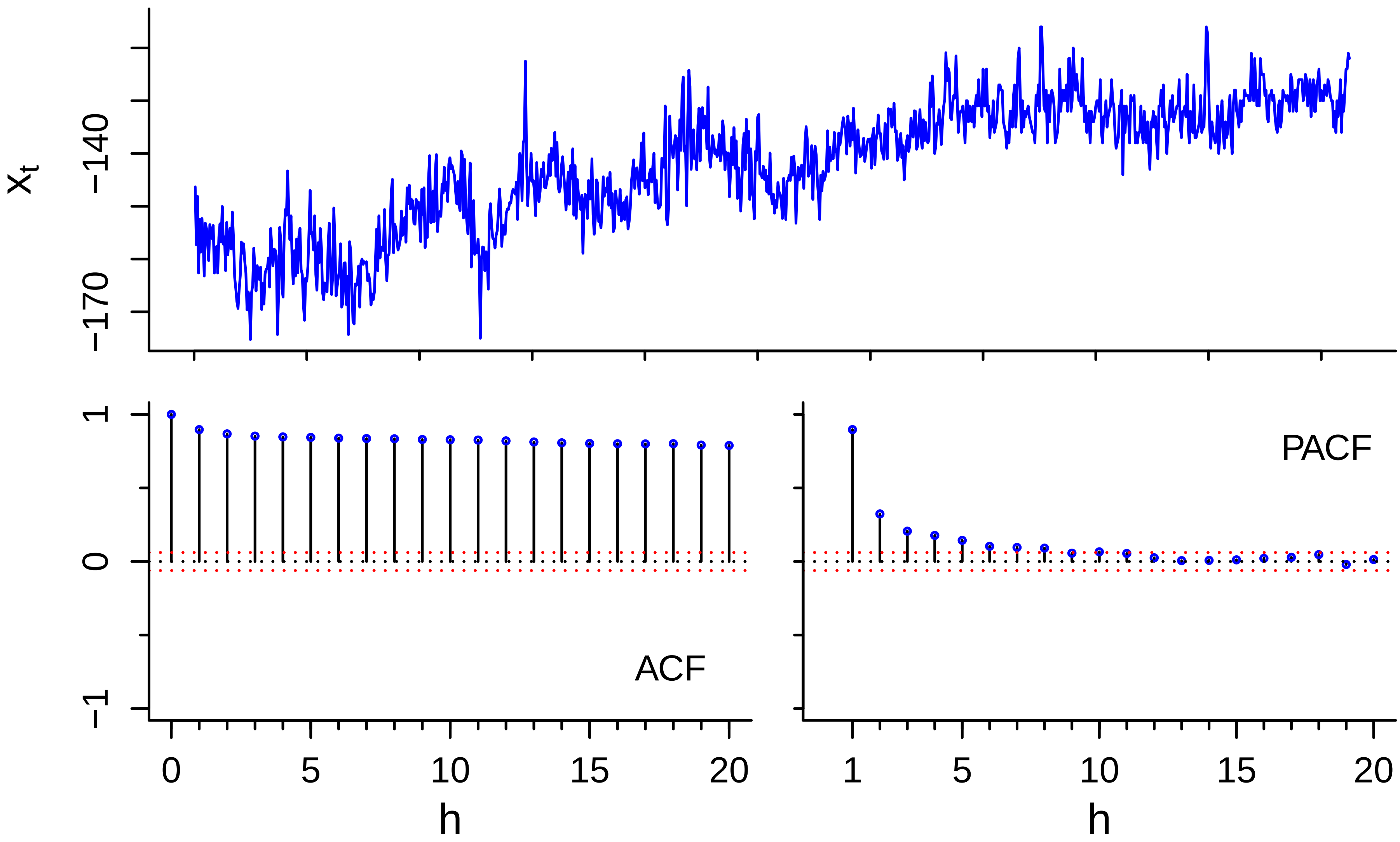


Diagnostic Tests for $\widehat{W}_j$

test	expected value	test statistic	p-value
turning point	440.67	413	**0.011**
difference-sign	331	335	0.591
rank	109726.5	116201	**0.023**
runs	332.1	302	**0.019**

AR method	AICC order	AIC order
Yule–Walker	0	0
Burg	0	0
OLS	0	**26**
MLE	0	0

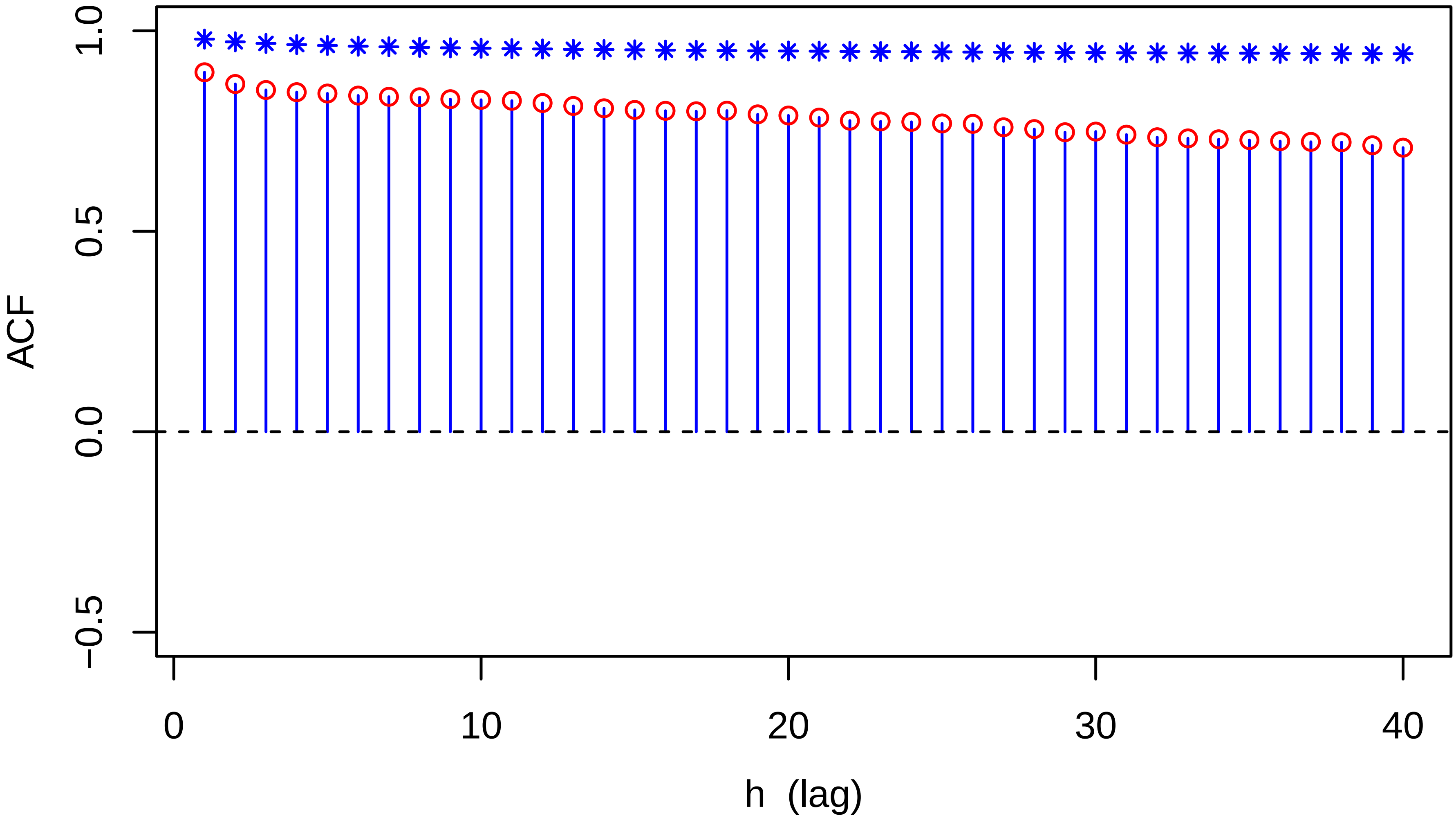
1st Difference of Atomic Clock Time Series



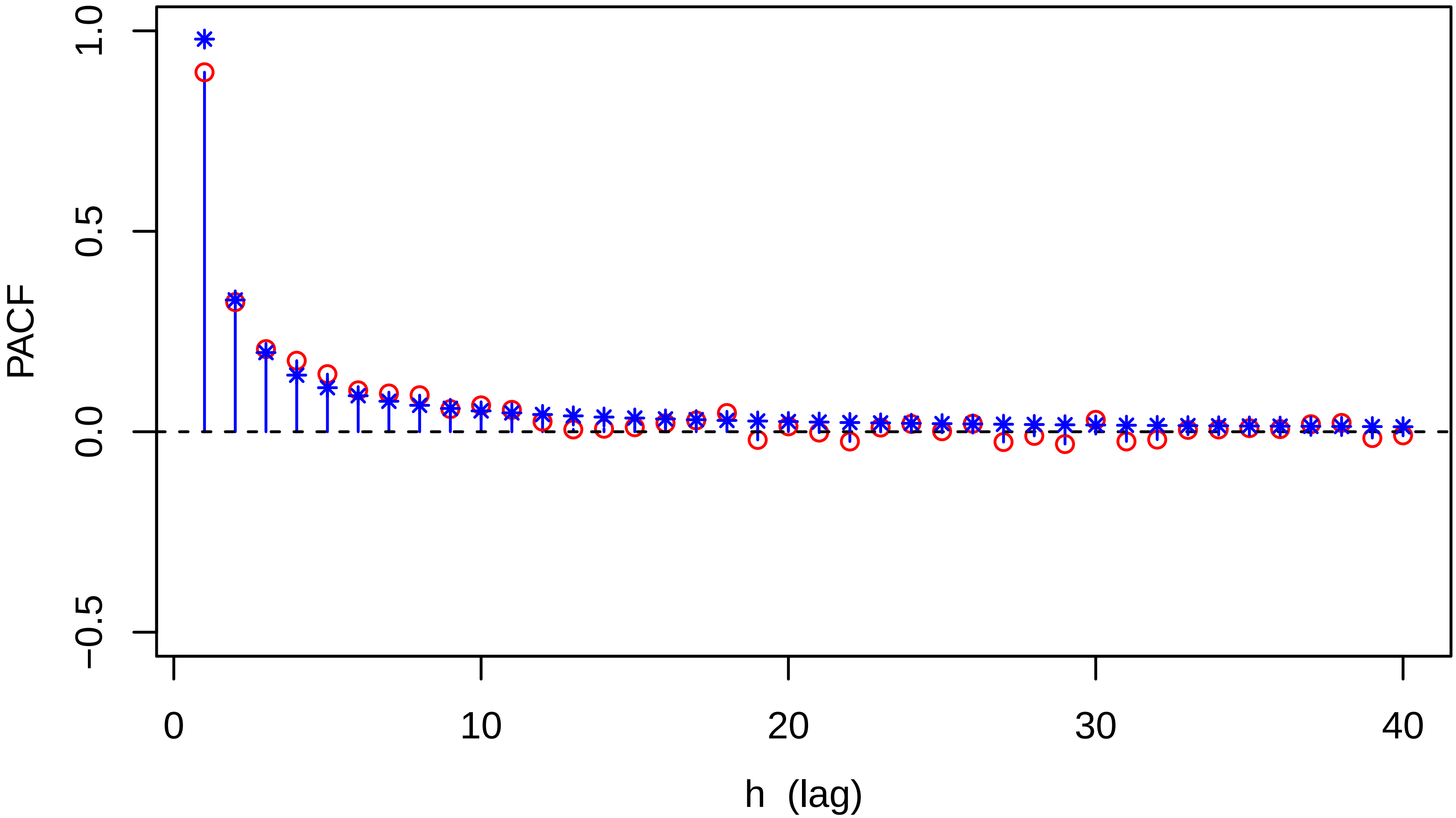
Atomic Clock 1st Difference as FD Process

- as second example, consider first difference ∇X_t of atomic clock time measurements
- using `fracdiff` to fit FD(δ) yields $\hat{\delta} \doteq 0.495$
- fact that $\hat{\delta}$ is so close to stationary/nonstationary boundary 1/2 cause for concern (particularly since `fracdiff` forces constraint $0 \leq \delta < 1/2$)

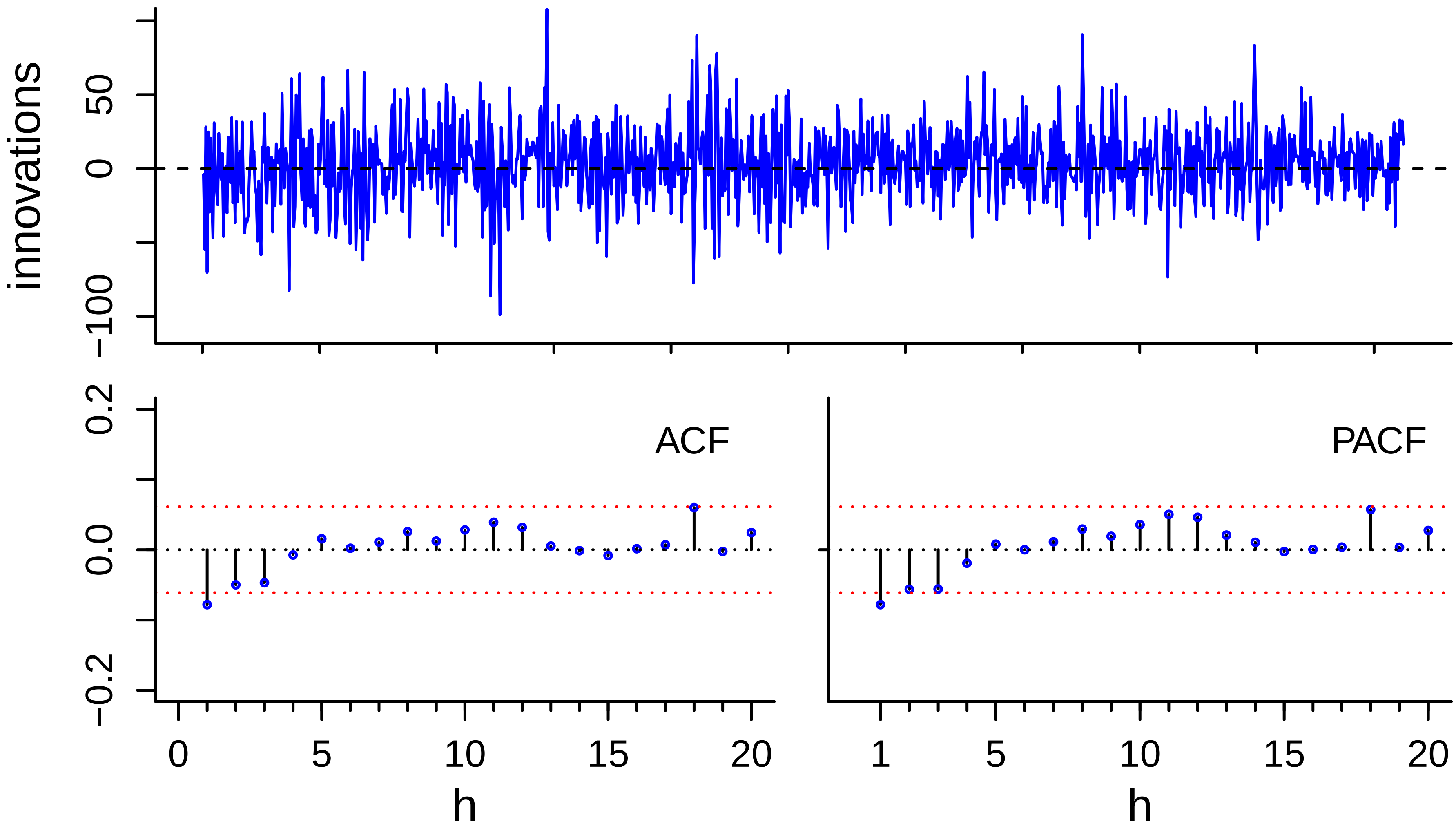
Sample & FD ACF for Atomic Clock ∇X_t



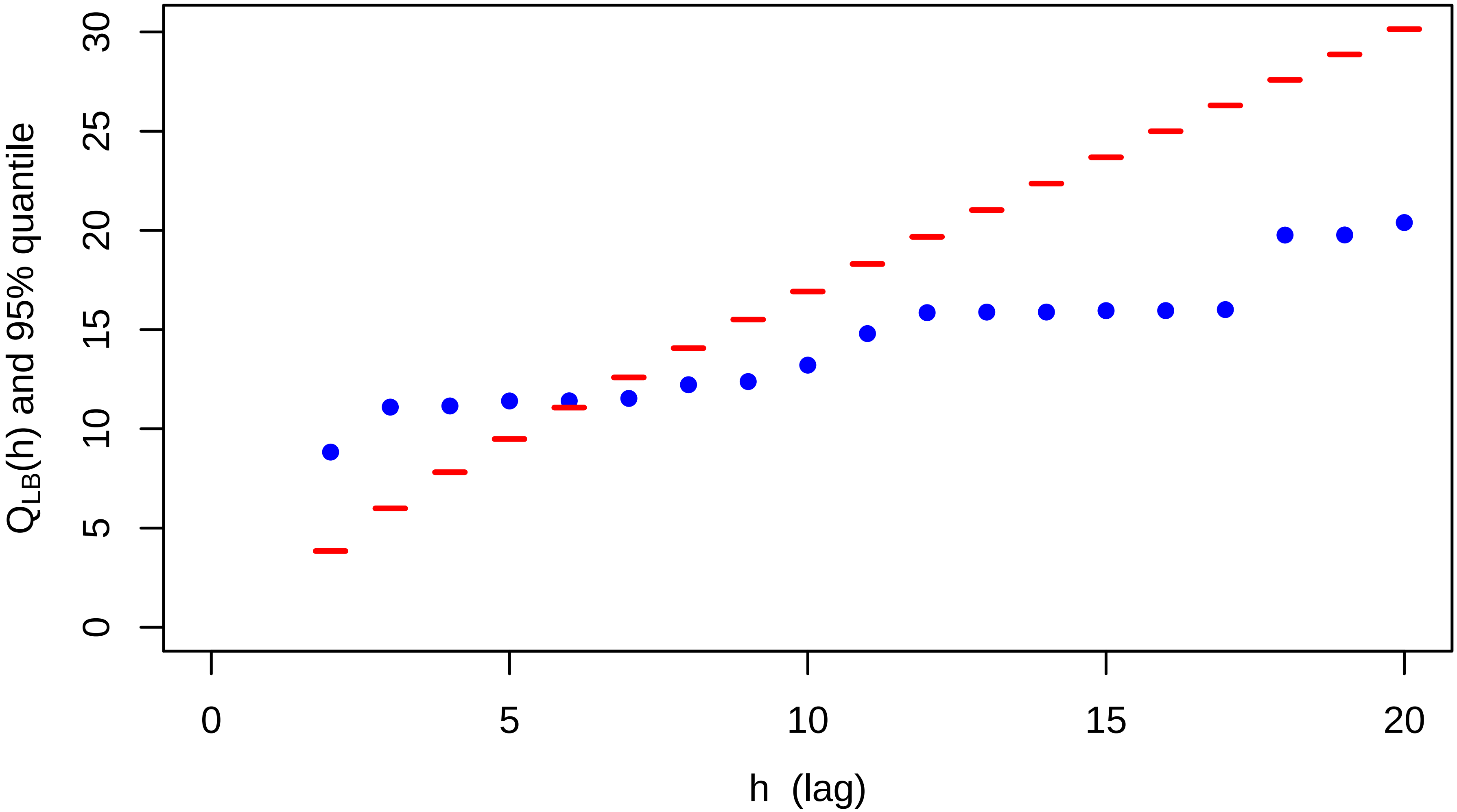
Sample & FD PACF for Atomic Clock ∇X_t



Normalized Innovations for $\widehat{W}_j$



Portmanteau Tests for $\widehat{W}_j$



Diagnostic Tests for $\widehat{W}_j$

test	expected value	test statistic	p-value
turning point	682	704	0.103
difference-sign	512	507	0.589
rank	262400	271787	0.086
runs	506.4	538	**0.045**

AR method	AICC order	AIC order
Yule–Walker	**4**	**3**
Burg	**4**	**3**
OLS	**4**	**4**
MLE	**4**	**3**

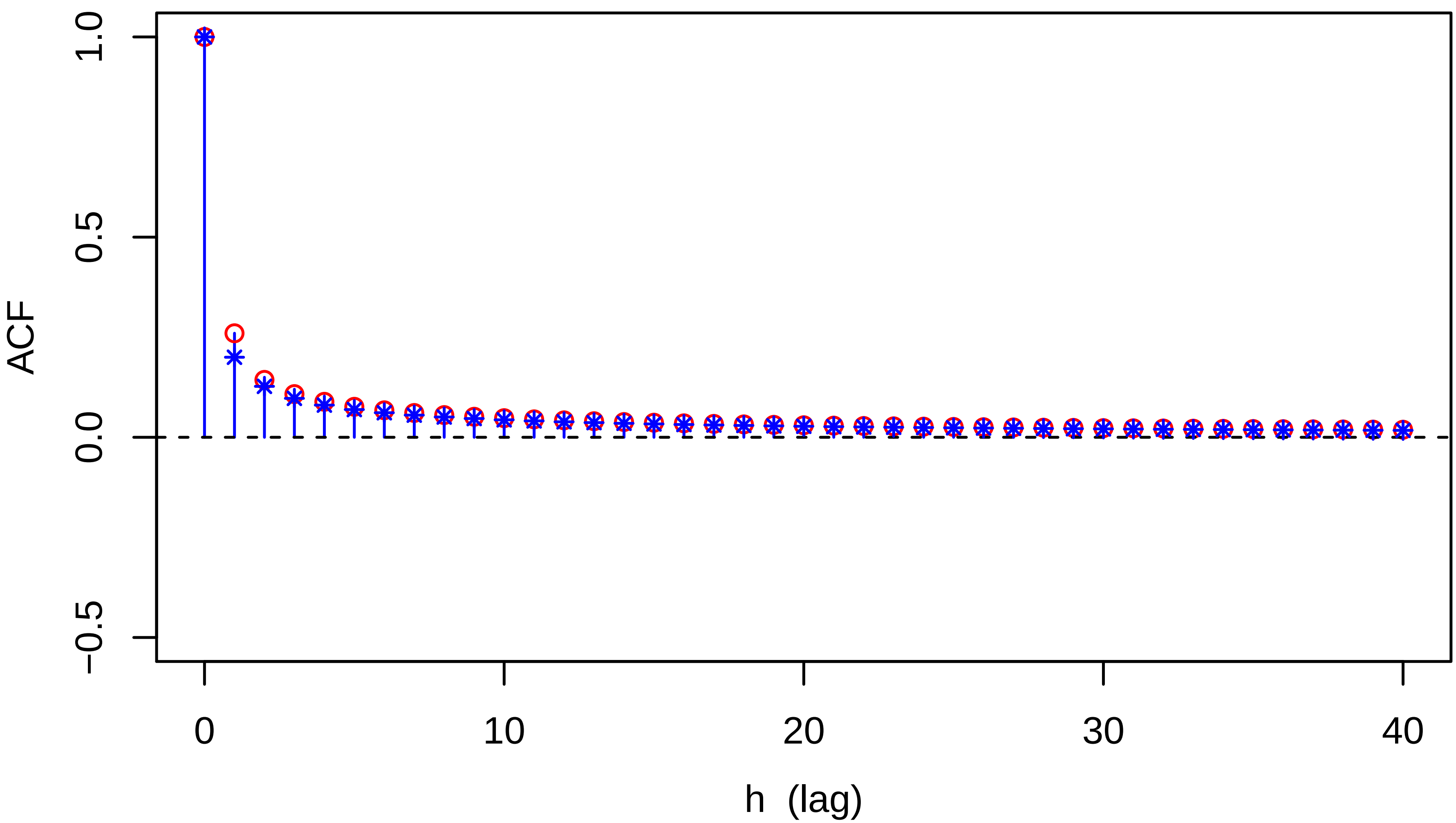
Fractional Gaussian Noise: I

- FD processes are not the only option for modeling time series with long-range dependence
- model predating FD process is fractional Gaussian noise (FGN), proposed by Mandelbrot & van Ness (1968)
- FGN is formed from increments of fractional Brownian motion, which is a continuous parameter process exhibiting so-called self-similarity properties (related to notion of fractals)
- for our purposes, can regard FGN as a Gaussian stationary process X_t with zero mean and with an ACVF given by
$$\gamma(h) = \frac{\sigma_X^2}{2}\left(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H}\right),$$
where σ_X^2 is the process variance, and H is the so-called Hurst (or self-similarity) parameter satisfying $0 < H < 1$

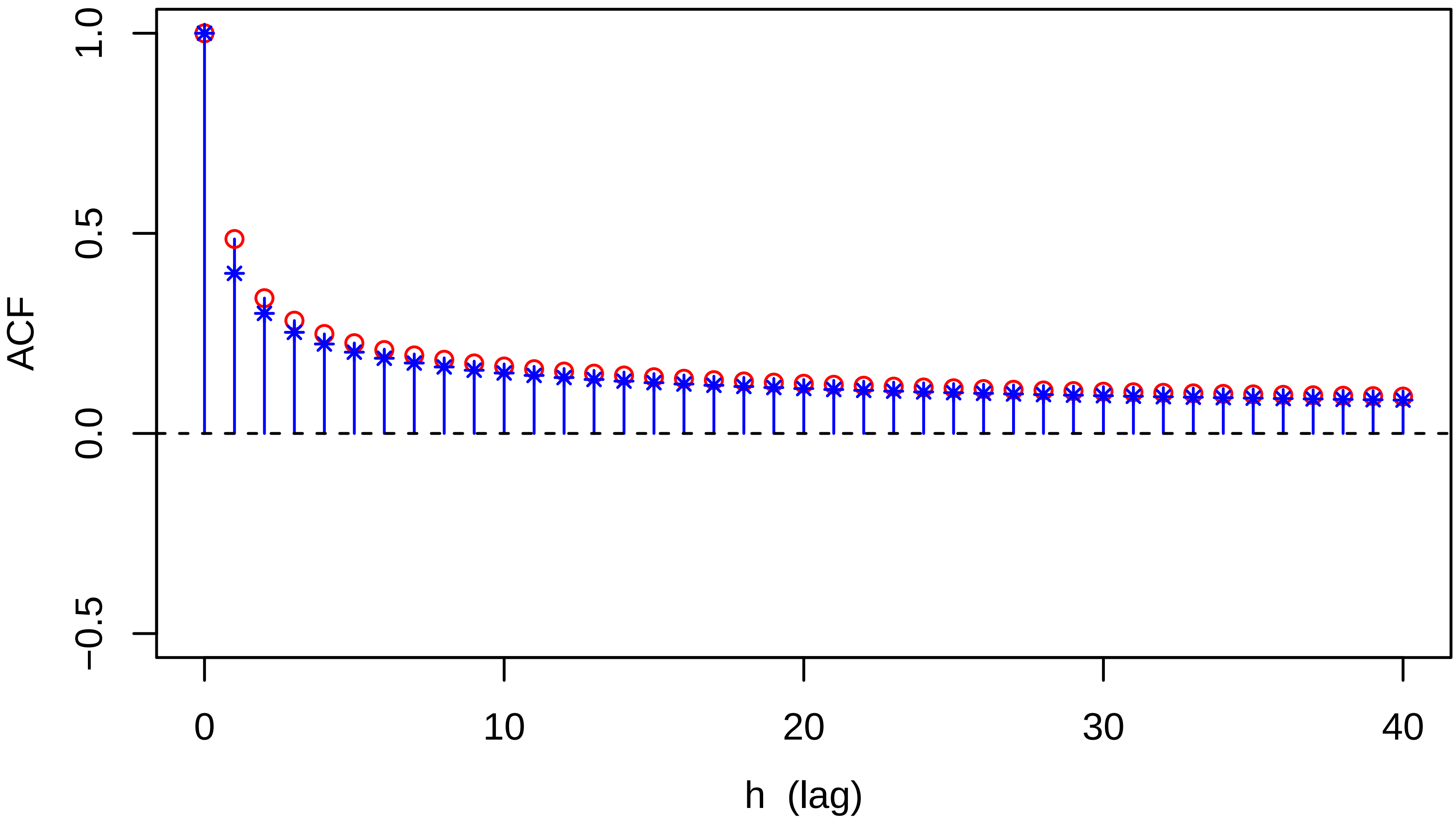
Fractional Gaussian Noise: II

- FD processes and FGNs are remarkably similar models if we make the correspondence $\delta = H - \frac{1}{2}$
- following overheads show
  - ACFs and PACFs for FD process and FGN for four settings of δ considered earlier (0.167, 0.286, 0.444 and 0.474), with H set to $\delta + \frac{1}{2}$
  - exact simulations of paired FD/FGN processes using circulant embedding with the same set of random deviates used to create all realizations

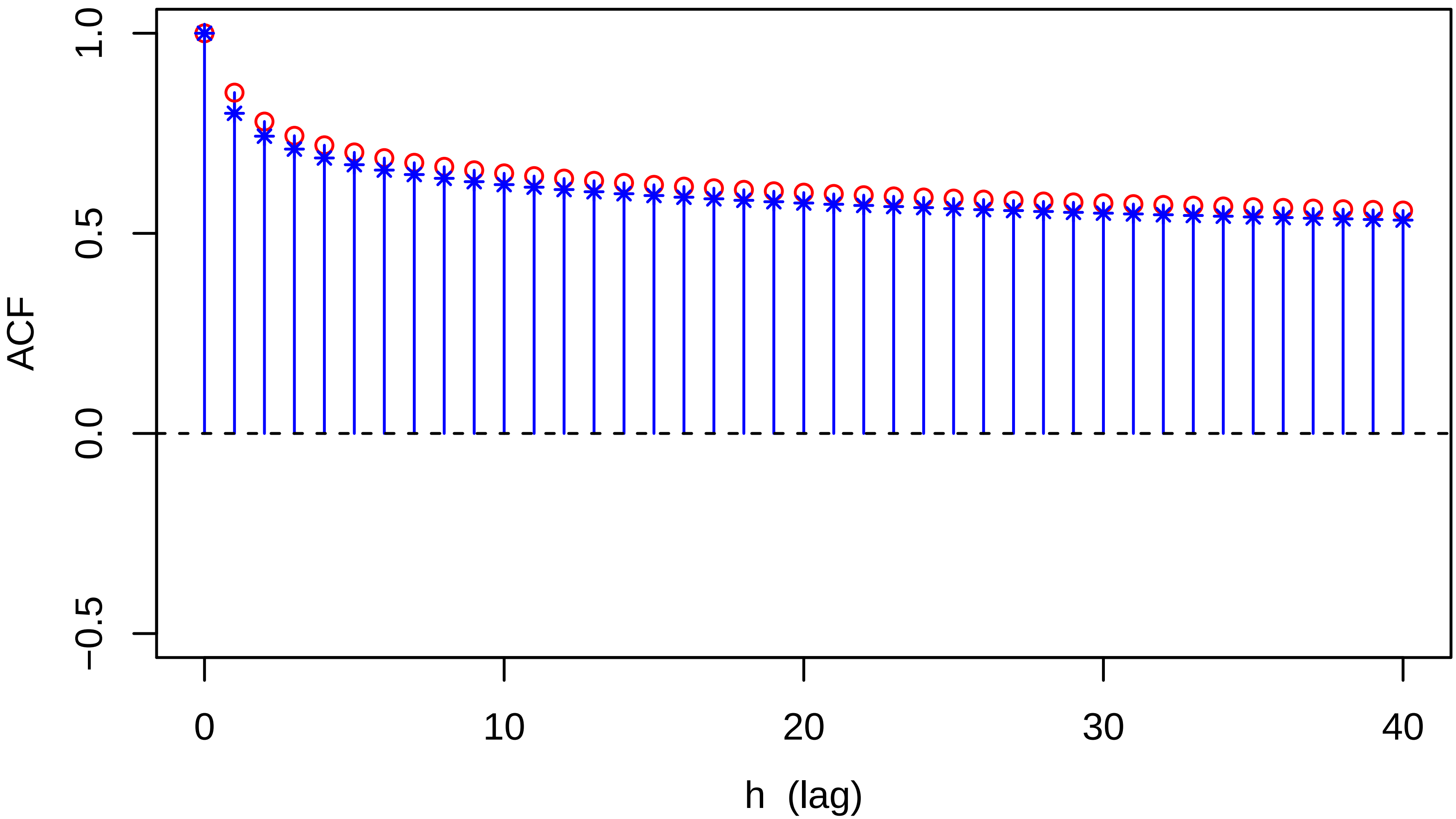
Comparison of FGN (∘) and FD (∗) ACFs, $\delta \doteq 0.167$



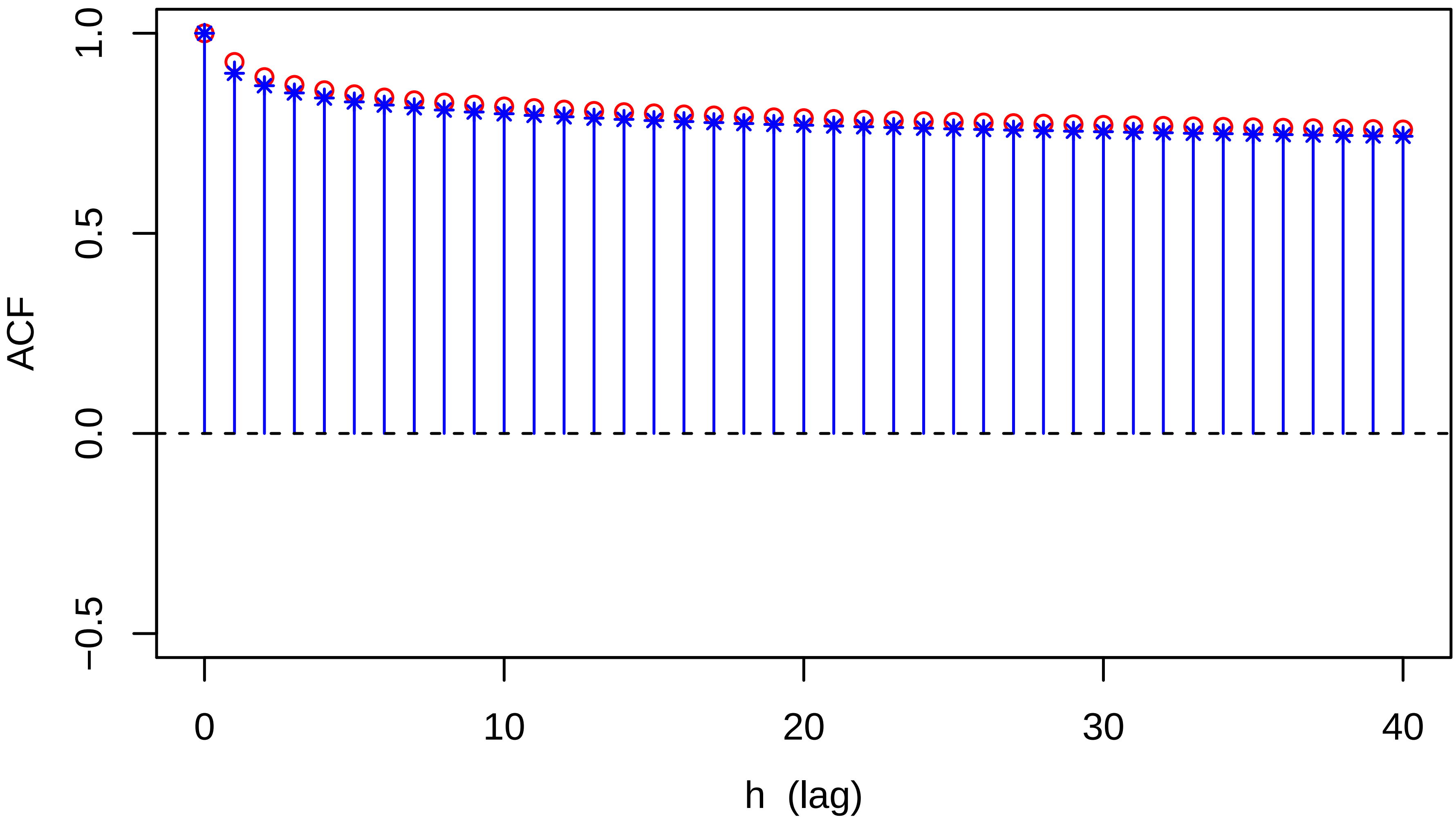
Comparison of FGN (∘) and FD (∗) ACFs, $\delta \doteq 0.286$



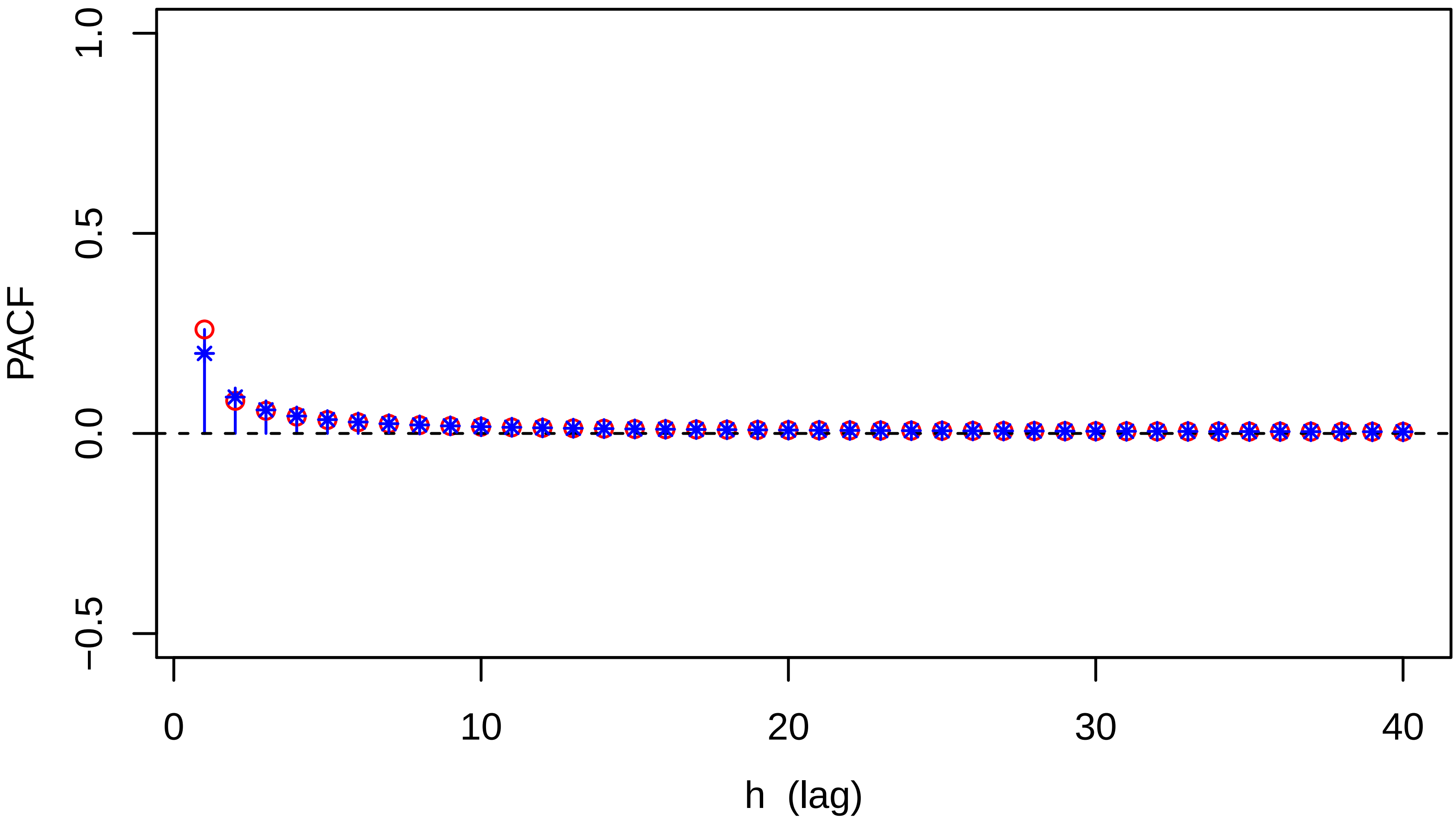
Comparison of FGN (∘) and FD (∗) ACFs, $\delta \doteq 0.444$



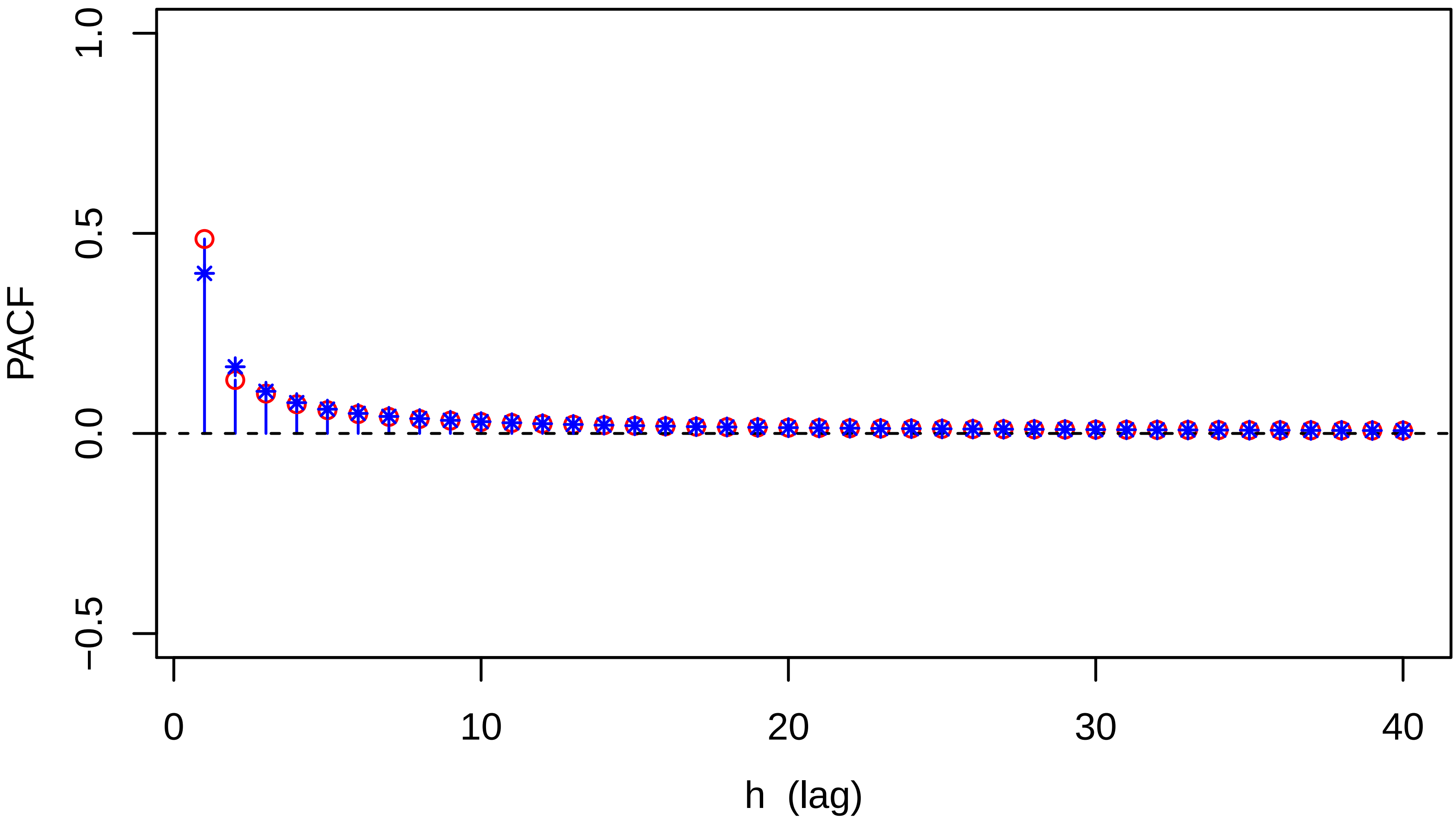
Comparison of FGN (∘) and FD (∗) ACFs, $\delta \doteq 0.474$



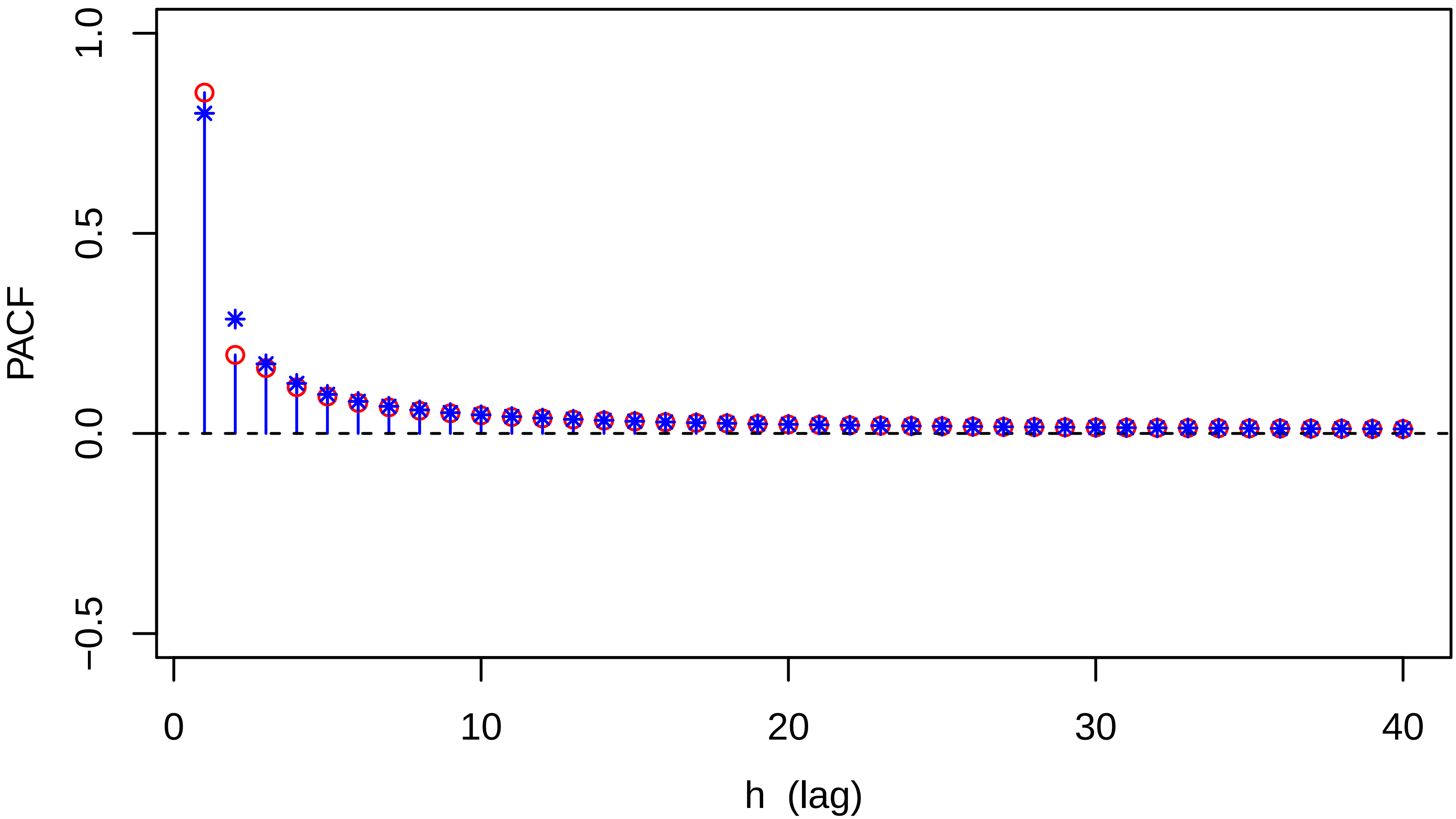
Comparison of FGN (∘) and FD (∗) PACFs, $\delta \doteq 0.167$



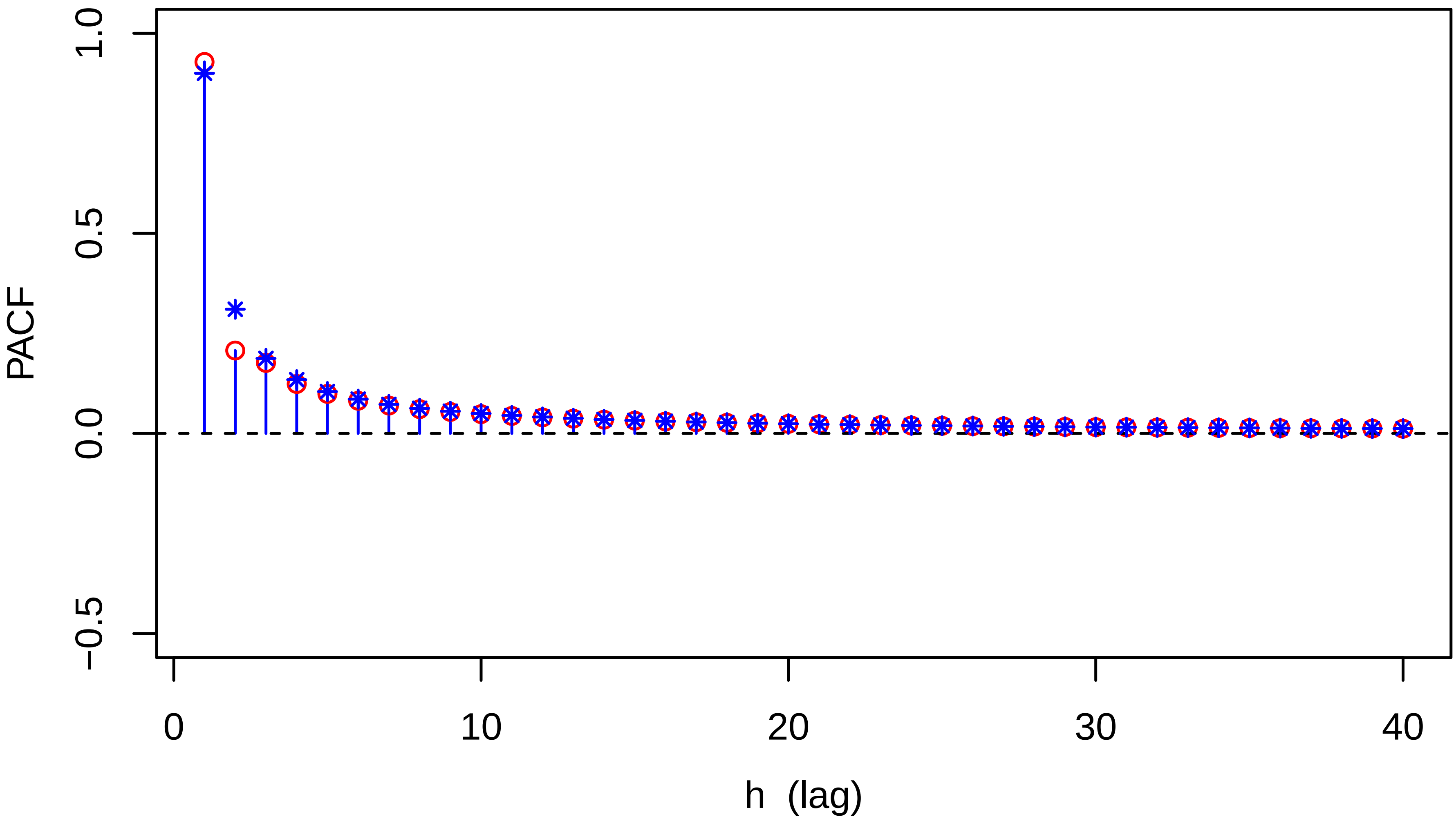
Comparison of FGN (∘) and FD (∗) PACFs, $\delta \doteq 0.286$



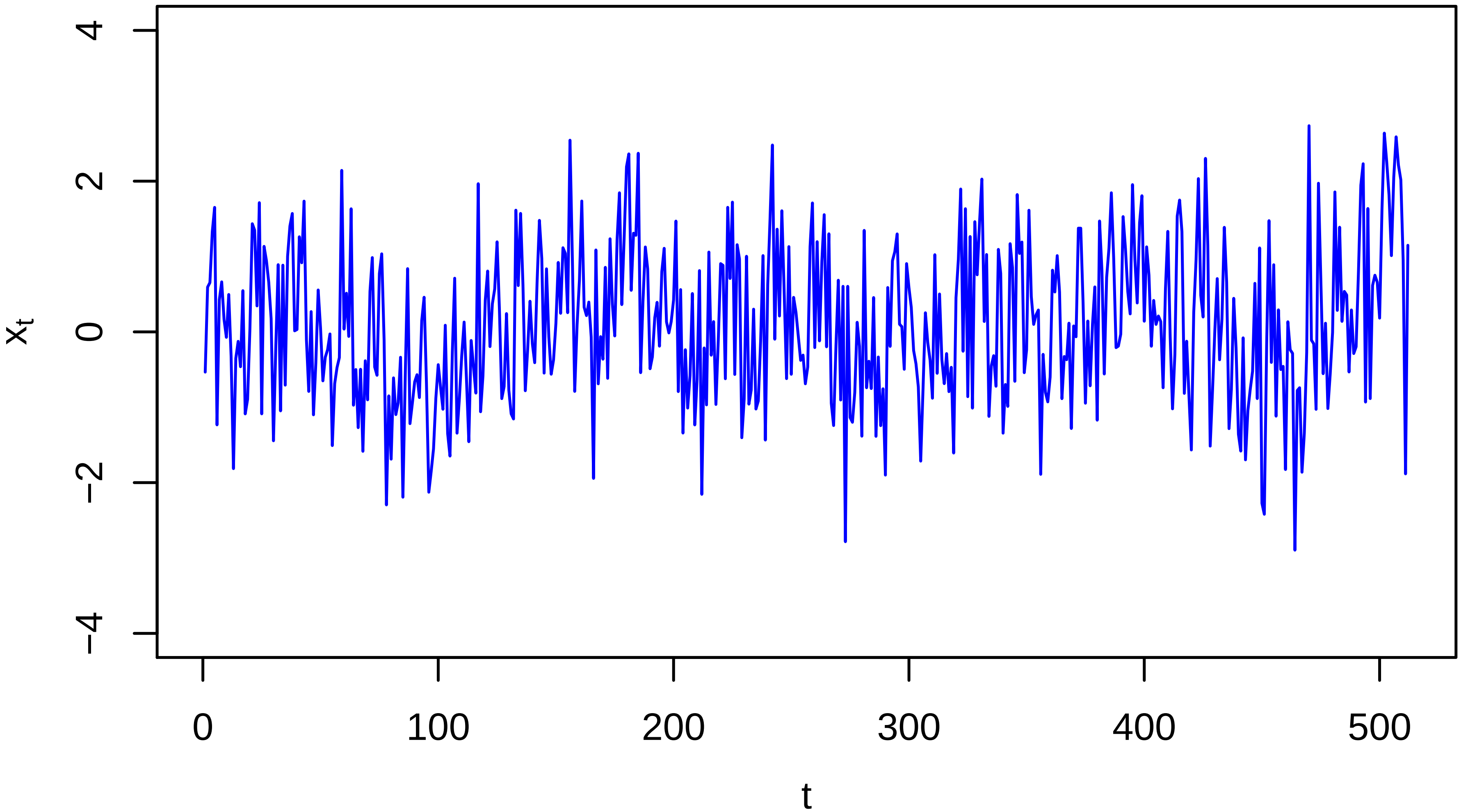
Comparison of FGN (∘) and FD (∗) PACFs, $\delta \doteq 0.444$



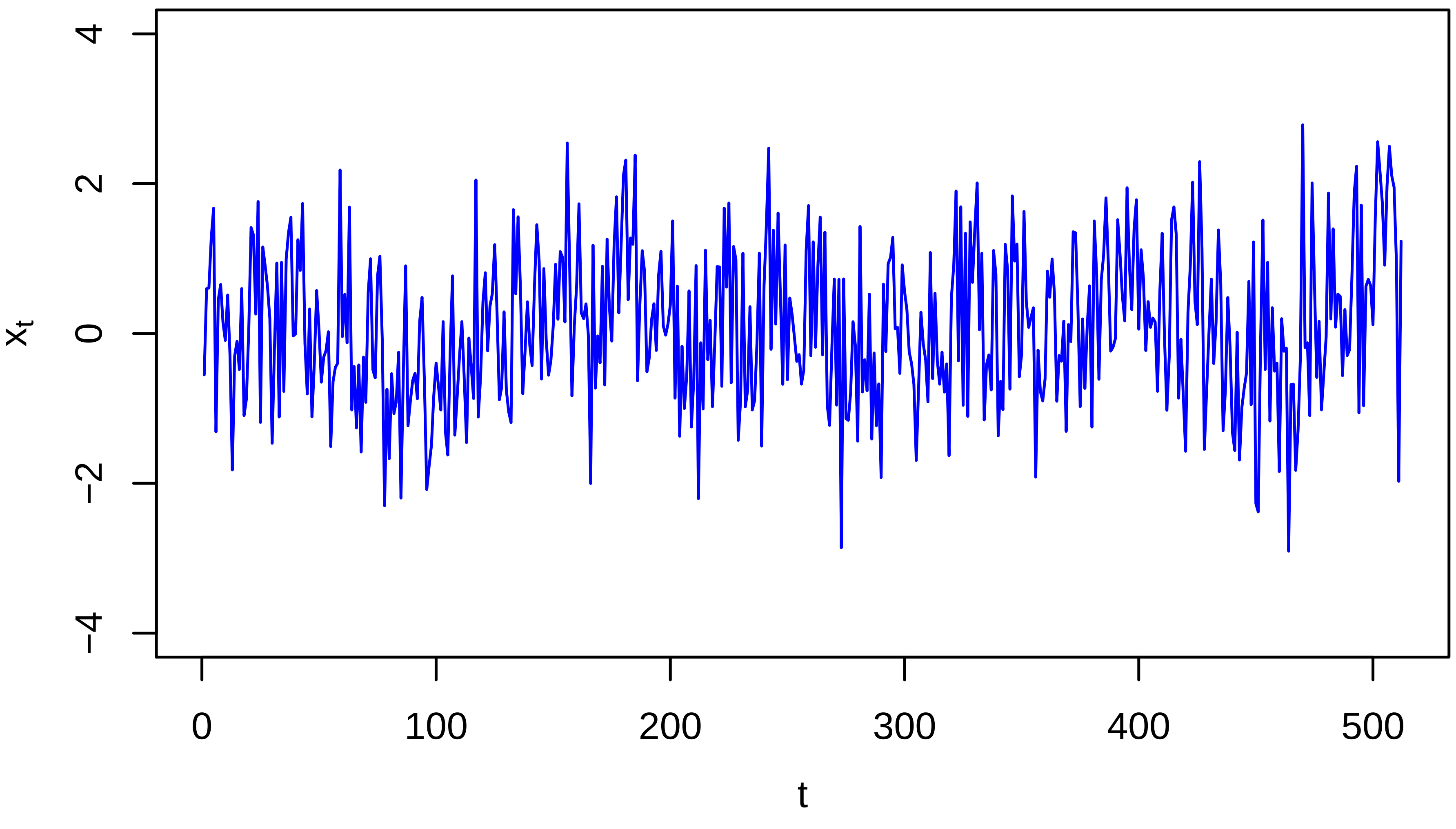
Comparison of FGN (∘) and FD (∗) PACFs, $\delta \doteq 0.474$



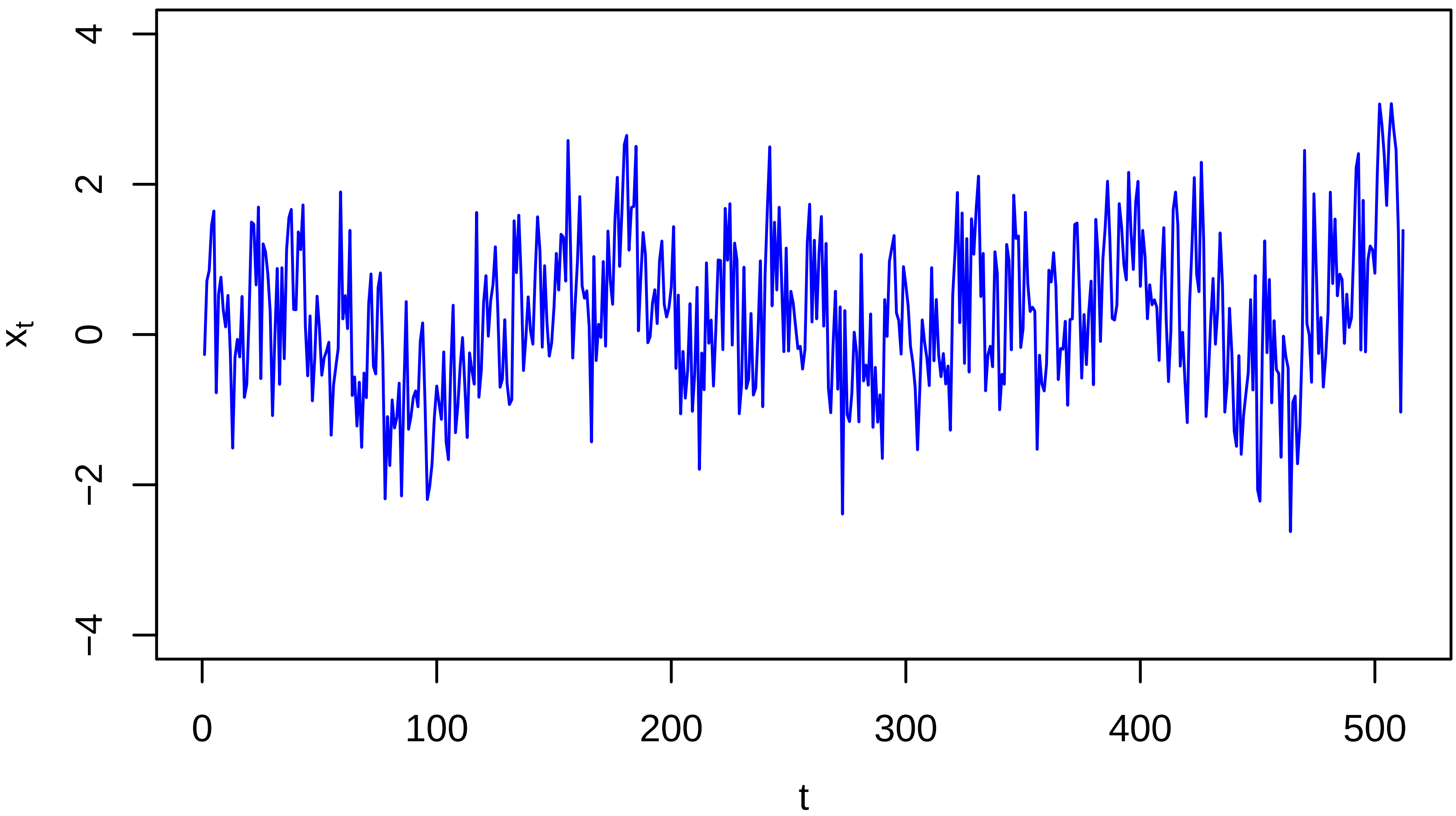
Simulated FGN Time Series, $H \doteq 0.667$



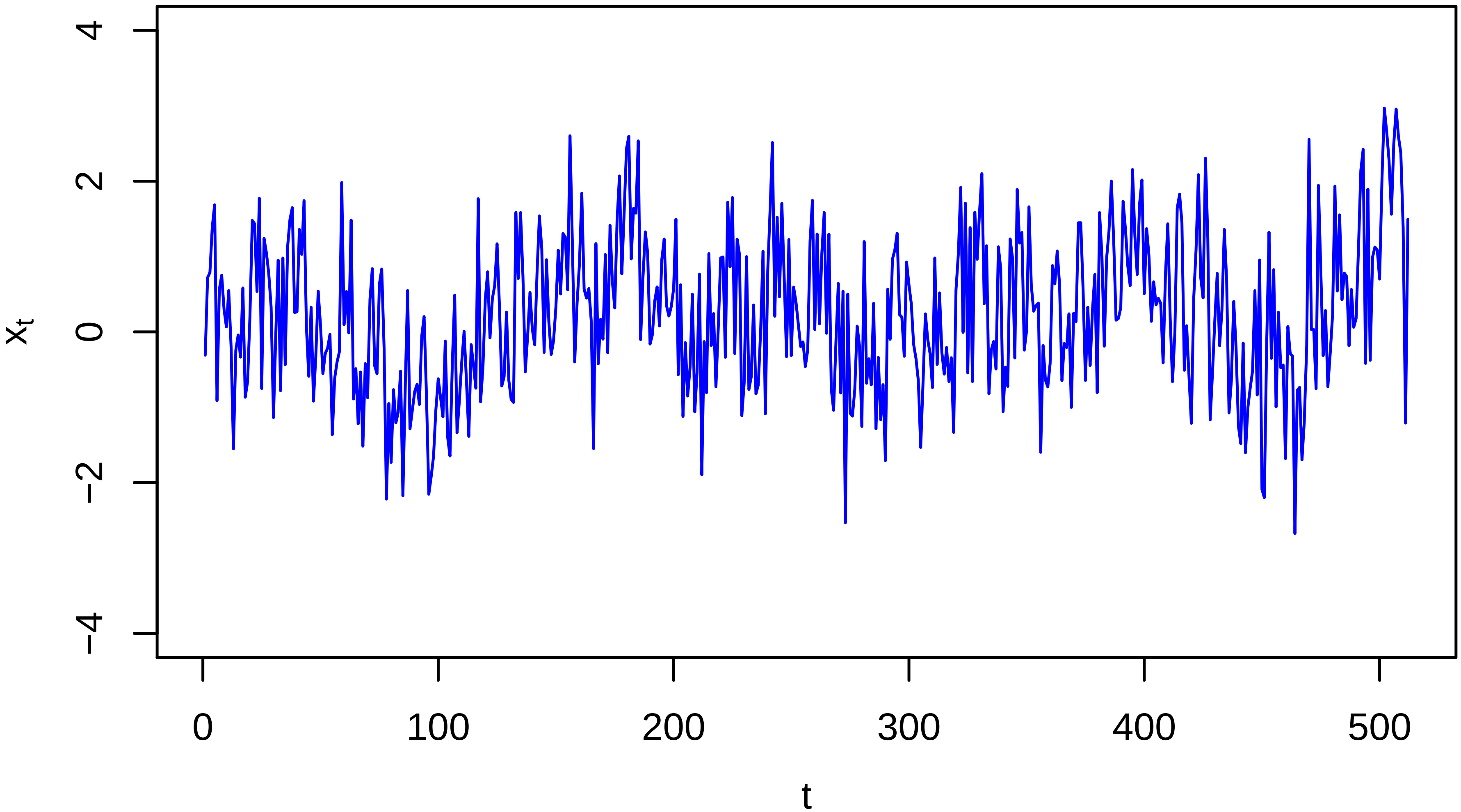
Simulated FD Time Series, $\delta \doteq 0.167$



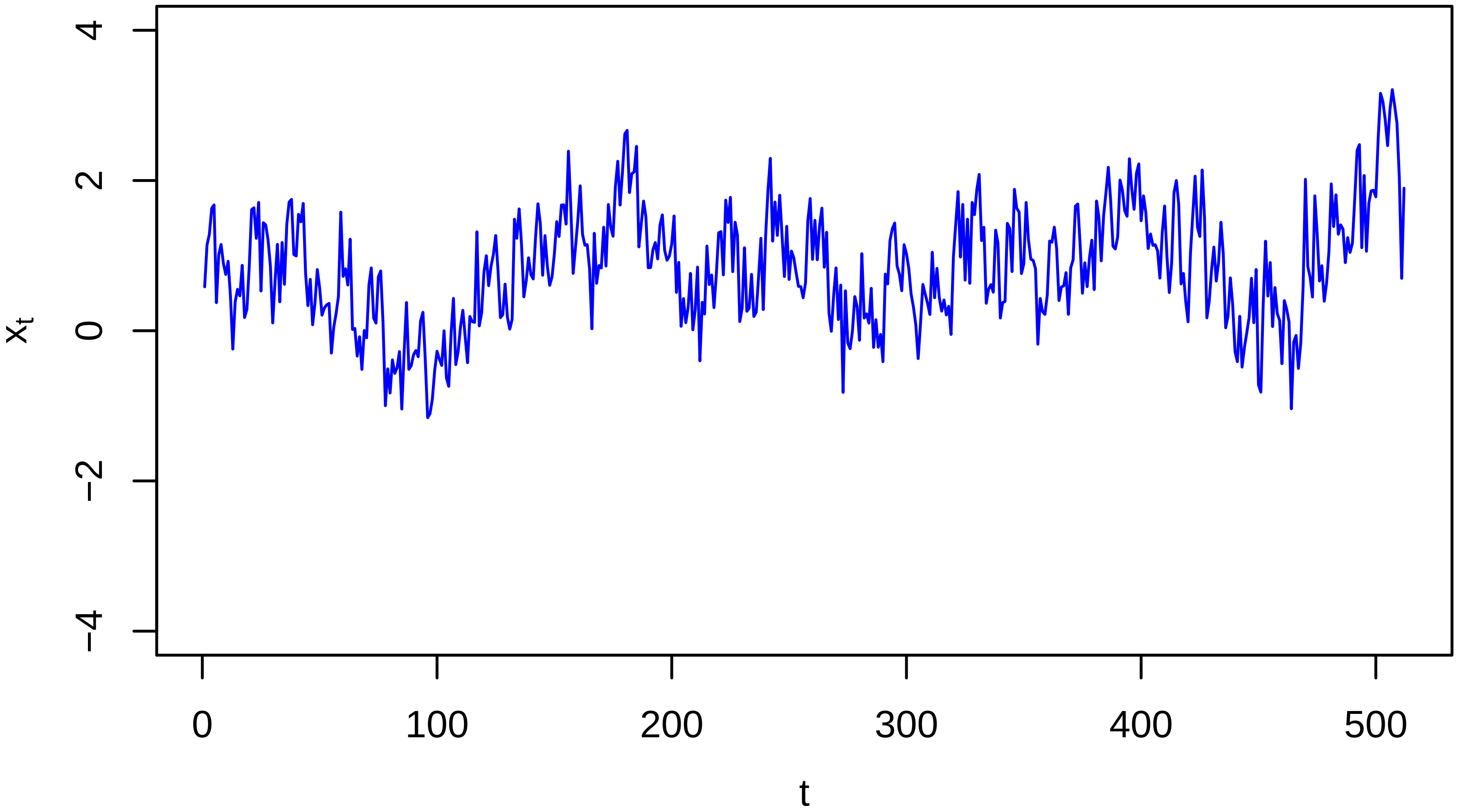
Simulated FGN Time Series, $H \doteq 0.786$



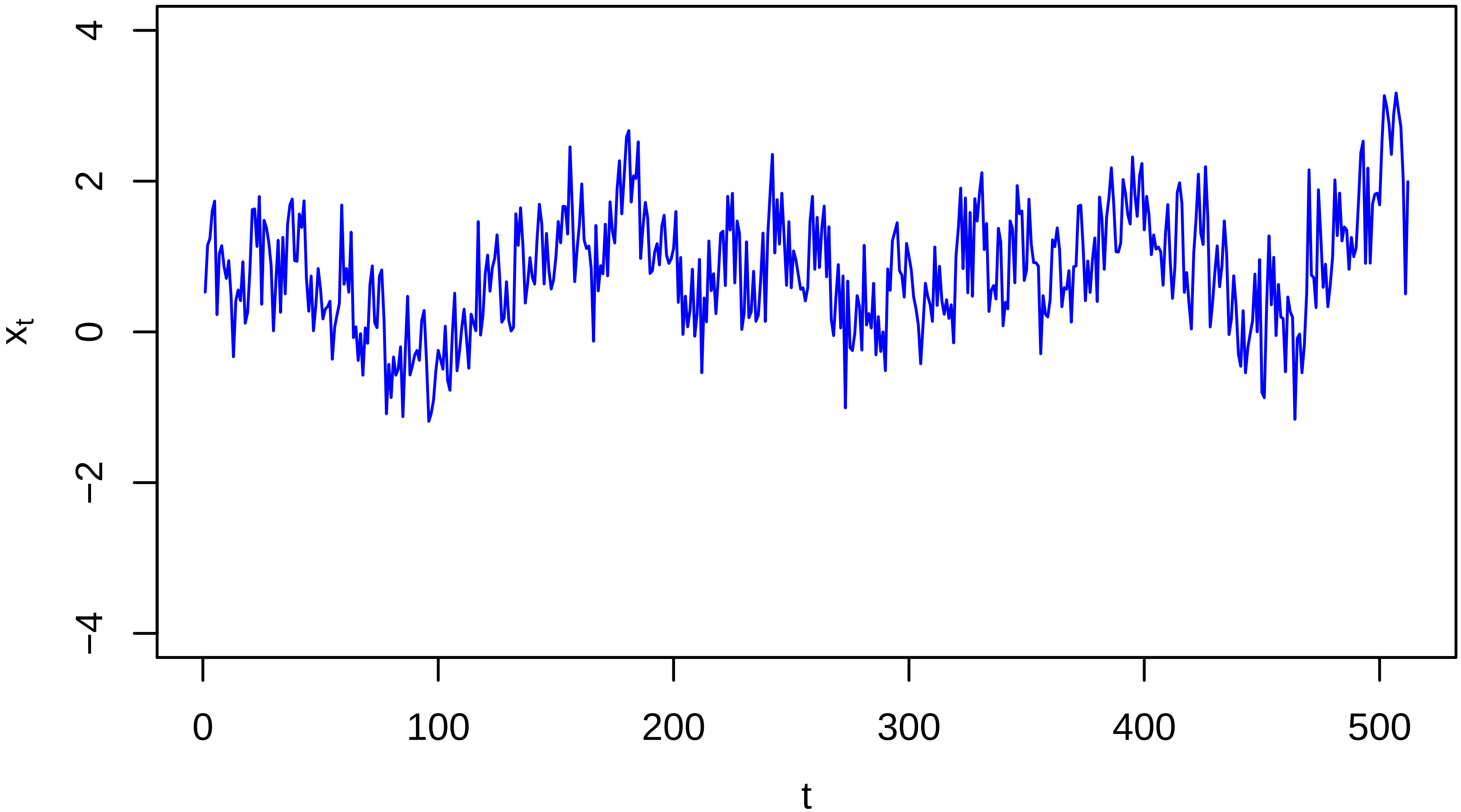
Simulated FD Time Series, $\delta \doteq 0.286$



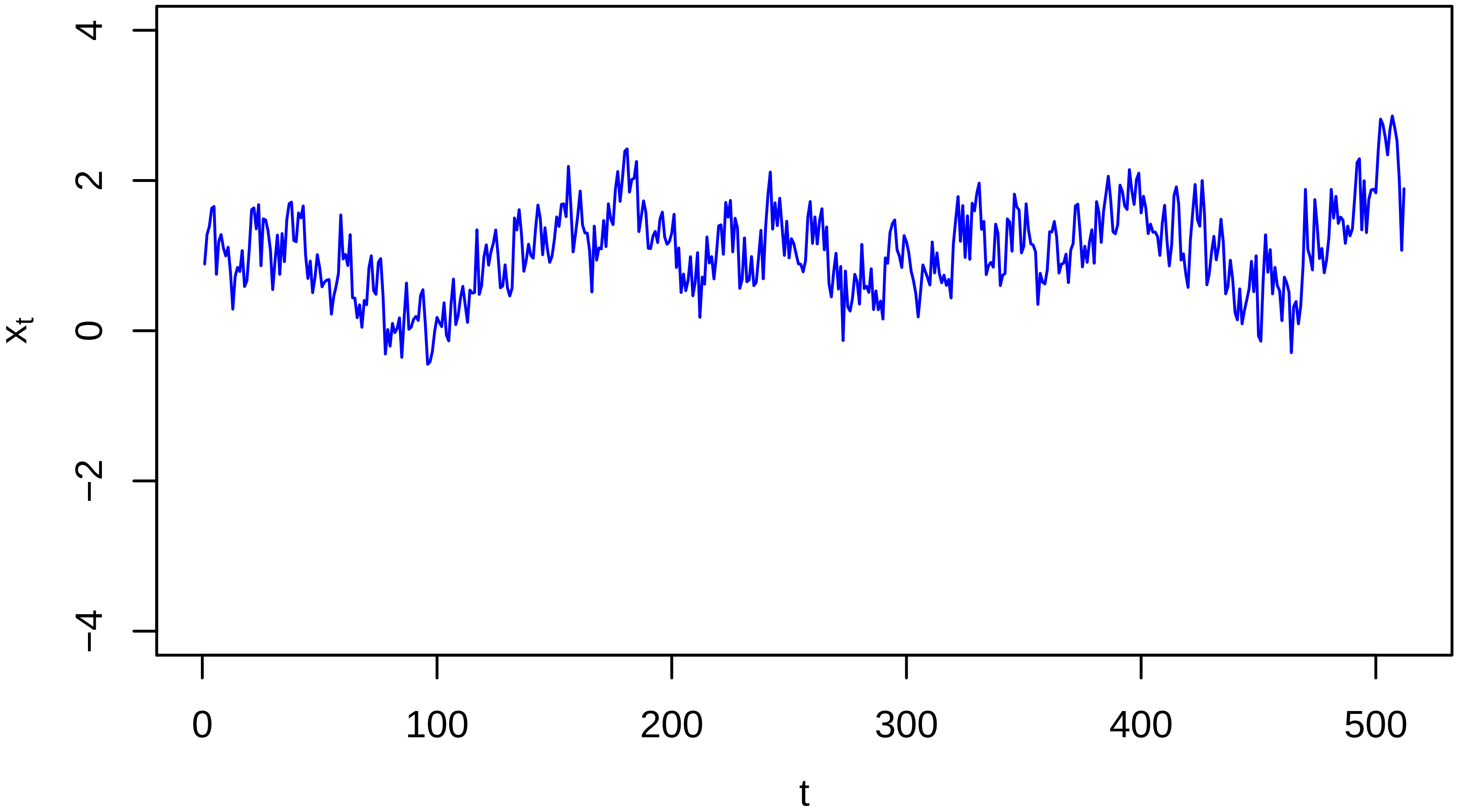
Simulated FGN Time Series, $H \doteq 0.944$



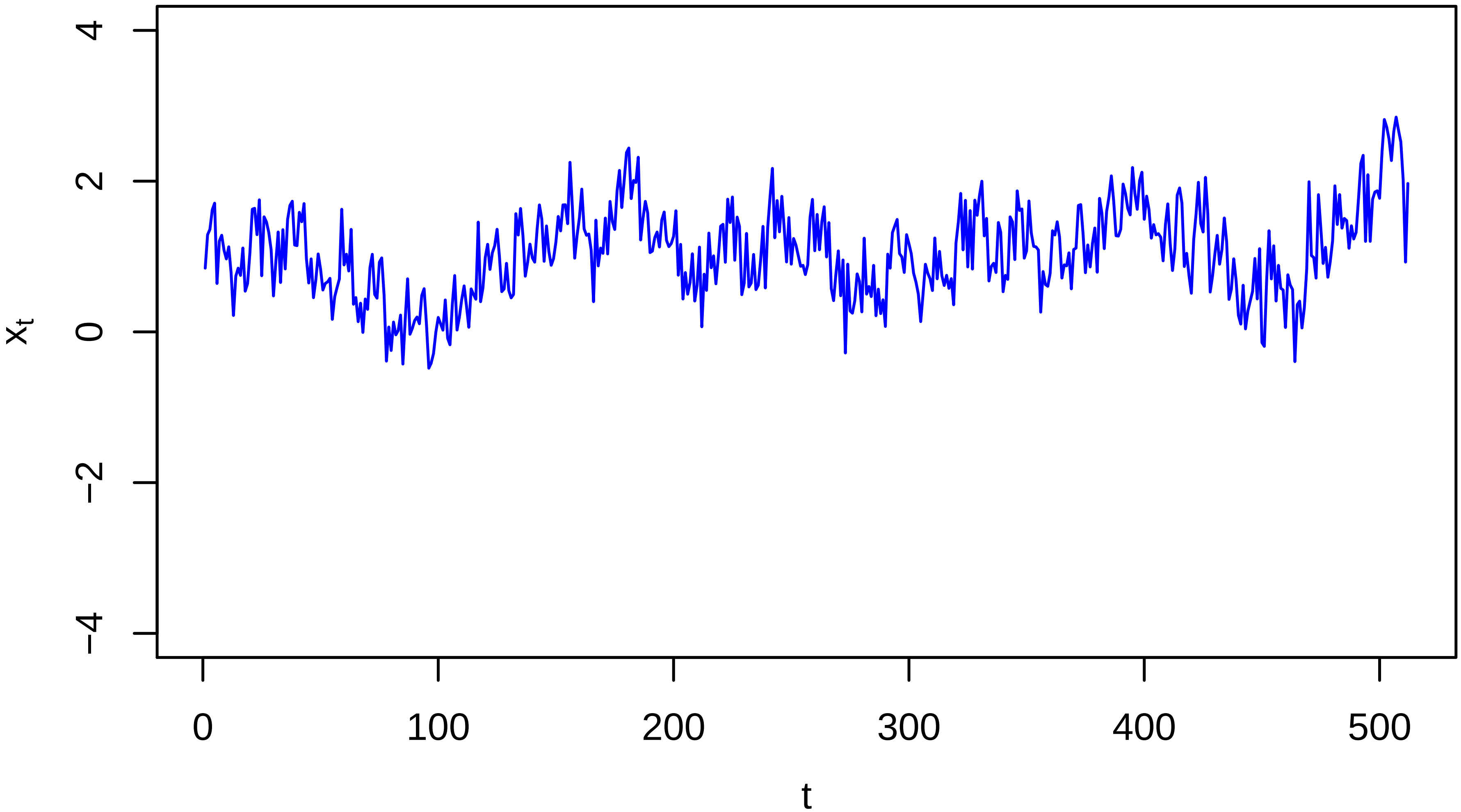
Simulated FD Time Series, $\delta \doteq 0.444$



Simulated FGN Time Series, $H \doteq 0.974$



Simulated FD Time Series, $\delta \doteq 0.474$



Fractional Gaussian Noise: III

- despite their similarities, FD processes X_t and FGNs Y_t have analytical properties with interesting differences
  - both have ACVFs that can be expressed in a simple manner:
$$\gamma_X(h) = \sigma^2 \frac{\sin(\pi\delta)\Gamma(1-2\delta)\Gamma(h+\delta)}{\pi\Gamma(h-\delta+1)}$$
$$\gamma_Y(h) = \frac{\sigma_Y^2}{2}\left(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H}\right)$$
    - if your tolerance for the Γ function is low, can use
$$\gamma_X(0) = \sigma^2 \frac{\Gamma(1-2\delta)}{\Gamma^2(1-\delta)} \text{ with } \gamma_X(h) = \gamma_X(h-1)\frac{h+\delta-1}{h-\delta}, \quad h = 1, 2, \ldots$$
    - note that $\gamma_Y(0) = \sigma_Y^2$
  - PACF for FD process given by $\phi_{X,h,h} = \frac{\delta}{h-\delta}$, while PACF for FGN does not have a simple expression

Fractional Gaussian Noise: IV

– both processes can be expressed as infinite moving averages:

$$X_t = \sum_{k=0}^{\infty} \psi_{X,k} Z_{t-k} \text{ and } Y_t = \sum_{k=0}^{\infty} \psi_{Y,k} Z_{t-k}$$

with

$$\psi_{X,k} = \frac{\Gamma(k+\delta)}{\Gamma(k+1)\Gamma(\delta)} \text{ and } \psi_{Y,k} = \text{hmmm} \ldots$$

– both possess spectral density functions, but one for FD process X_t is far simpler:

$$S_X(f) = \frac{\sigma^2}{[4\sin^2(\pi f)]^\delta}$$

$$S_Y(f) = 4\,\sigma_Y^2 C_H \sin^2(\pi f) \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}$$

Fractional Gaussian Noise: V

  - first difference of FD process is another FD process; first difference of FGN is *not* another FGN
  - using AR and MA filters, can convert an FD(δ) process into an ARFIMA(p, δ, q); can also do the same with FGN, but this approach is not widely used (little competitive advantage over ARFIMA approach)

- on the whole, FD processes have much to recommend them over FGNs, which is why FD processes are more popular in the statistical community (connection of FGNs to self similarity makes them popular for mathematical study)

References

- J. Beran (1994), *Statistics for Long-Memory Processes*, New York: Chapman & Hall
- P. F. Craigmile (2003), 'Simulating a Class of Stationary Gaussian Processes using the Davies–Harte Algorithm, with Application to Long Memory Processes,' *Journal of Time Series Analysis*, , pp. 505–511
- H. A. David (1985), 'Bias of S^2 Under Dependence,' *The American Statistician*, **39**, p. 201
- R. B. Davies and D. S. Harte (1987), 'Tests for Hurst Effect,' *Biometrika*, **74**, pp. 95–101
- T. Gneiting (2000), 'Power-law Correlations, Related Models for Long-Range Dependence, and Their Simulation,' *Journal of Applied Probability*, **37**, pp. 1104–1109
- C. W. J. Granger and R. Joyeux (1980), 'An Introduction to Long-Memory Time Series Models and Fractional Differencing,' *Journal of Time Series Analysis*, **1**, pp. 15–29
- J. Haslett and A. E. Raftery (1989), 'Long-memory Dependence: Assessing Ireland's Wind Power Resource' with Discussion, *Applied Statistics*, **38**, pp. 1–50
- J. R. M. Hosking (1981), 'Fractional Differencing,' *Biometrika*, **68**, pp. 165–176
- B. B. Mandelbrot and J. W. van Ness (1968), 'Fractional Brownian Motions, Fractional Noises and Applications,' *SIAM Review*, **10**, pp. 422–437