Beyond ARMA Models: Nonstationarity & Seasonality

- now know how to handle time series in a certain 'comfort zone'
  - no worrisome trends or seasonal patterns
  - sample ACF and PACF decay fairly rapidly
  - simple ARMA model quantifies correlation structure ('simple' meaning $p + q$ is relatively small)
- can handle certain 'out of zone' time series through manipulations yielding series that are back in comfort zone
  1. differencing – leads to ARIMA models for handling certain types of nonstationary time series (next topic for discussion)
  2. seasonal differencing – leads to SARIMA models for handling certain types of seasonality
  3. regression analysis with ARMA errors – designed to handle certain types of trends

Intrinsically Stationary Processes

- stochastic process $\{X_t\}$ said to be intrinsically stationary of integer order $d > 0$ if $\{X_t\}$, $\{\nabla X_t\}$, ..., $\{\nabla^{d-1} X_t\}$ are non-stationary, but $\{\nabla^d X_t\}$ is a stationary process, where

$$\begin{aligned}
\nabla X_t &\stackrel{\text{def}}{=} (1-B)X_t = X_t - X_{t-1} \\
\nabla^2 X_t &\stackrel{\text{def}}{=} \nabla(\nabla X_t) = (1-B)^2 X_t \\
&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\
&= X_t - 2X_{t-1} + X_{t-2} \\
&\vdots \\
\nabla^d X_t &\stackrel{\text{def}}{=} (1-B)^d X_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}
\end{aligned}$$

- convenient to refer to stationary process as being intrinsically stationary of order $d = 0$

ARIMA(p, d, q) Processes

- process $\{X_t\}$ is an ARIMA(p, d, q) process if
  1. $\{X_t\}$ is intrinsically stationary of order d and
  2. $\{\nabla^d X_t\}$ is an ARMA(p, q) process
- with $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, can express model as

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

  or, equivalently, as

$$\phi^*(B) X_t = \theta(B) Z_t,$$

  where $\phi^*(B) = \phi(B)(1 - B)^d$
- example: for ARIMA(1,1,0) model, have

$$\phi^*(B) = (1 - \phi B)(1 - B) = 1 - (1 + \phi)B + \phi B^2$$

ARIMA(0,1,0) Process: I

- simplest example of ARIMA process is ARIMA(0,1,0):

$$(1 - B)X_t = X_t - X_{t-1} = Z_t,$$

for which, assuming existence of X_0 and assuming $t \geq 1$,

$$\begin{aligned} X_1 &= X_0 + Z_1 \\ X_2 &= X_1 + Z_2 = X_0 + Z_1 + Z_2 \\ X_3 &= X_2 + Z_3 = X_0 + Z_1 + Z_2 + Z_3 \\ &\vdots \\ X_t &= X_0 + \sum_{u=1}^{t} Z_u, \end{aligned}$$

- above is a random walk starting from X_0

ARIMA(0,1,0) Process: II

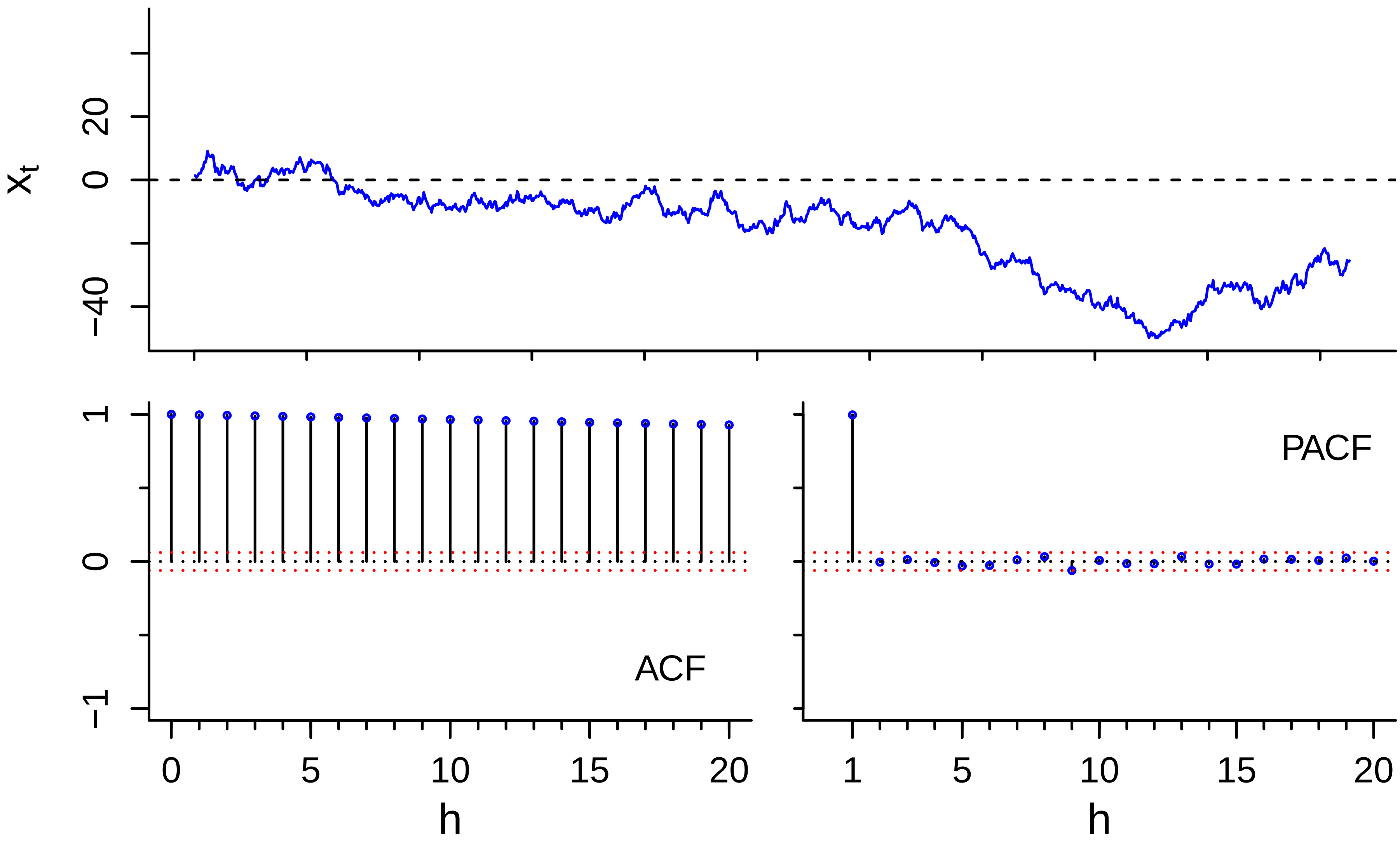
- assuming RV X_0 is uncorrelated with Z_t's, have

$$\begin{aligned}\operatorname{var}\{X_t\} &= \operatorname{var}\left\{X_0 + \sum_{u=1}^{t} Z_u\right\} \\ &= \operatorname{var}\{X_0\} + \sum_{u=1}^{t} \operatorname{var}\{Z_u\} \\ &= \operatorname{var}\{X_0\} + t\sigma^2,\end{aligned}$$

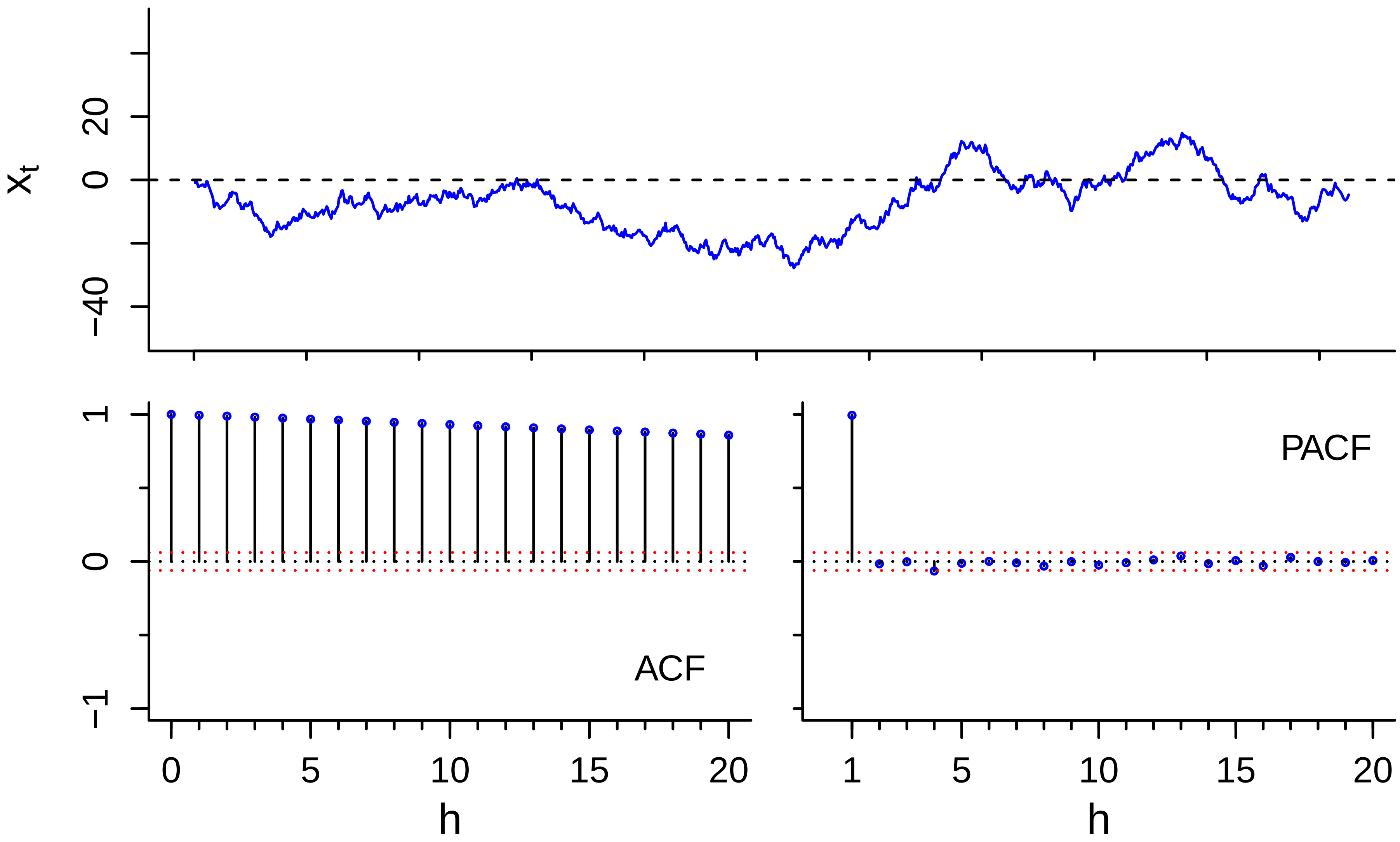
  which is either time-dependent or infinite (if $\operatorname{var}\{X_0\} = \infty$)

- ARIMA(0,1,0) is thus a nonstationary process (same true for all ARIMA(p, d, q) processes when d is a positive integer)
- let's look at three realizations of a random walk ($n = 1026$)

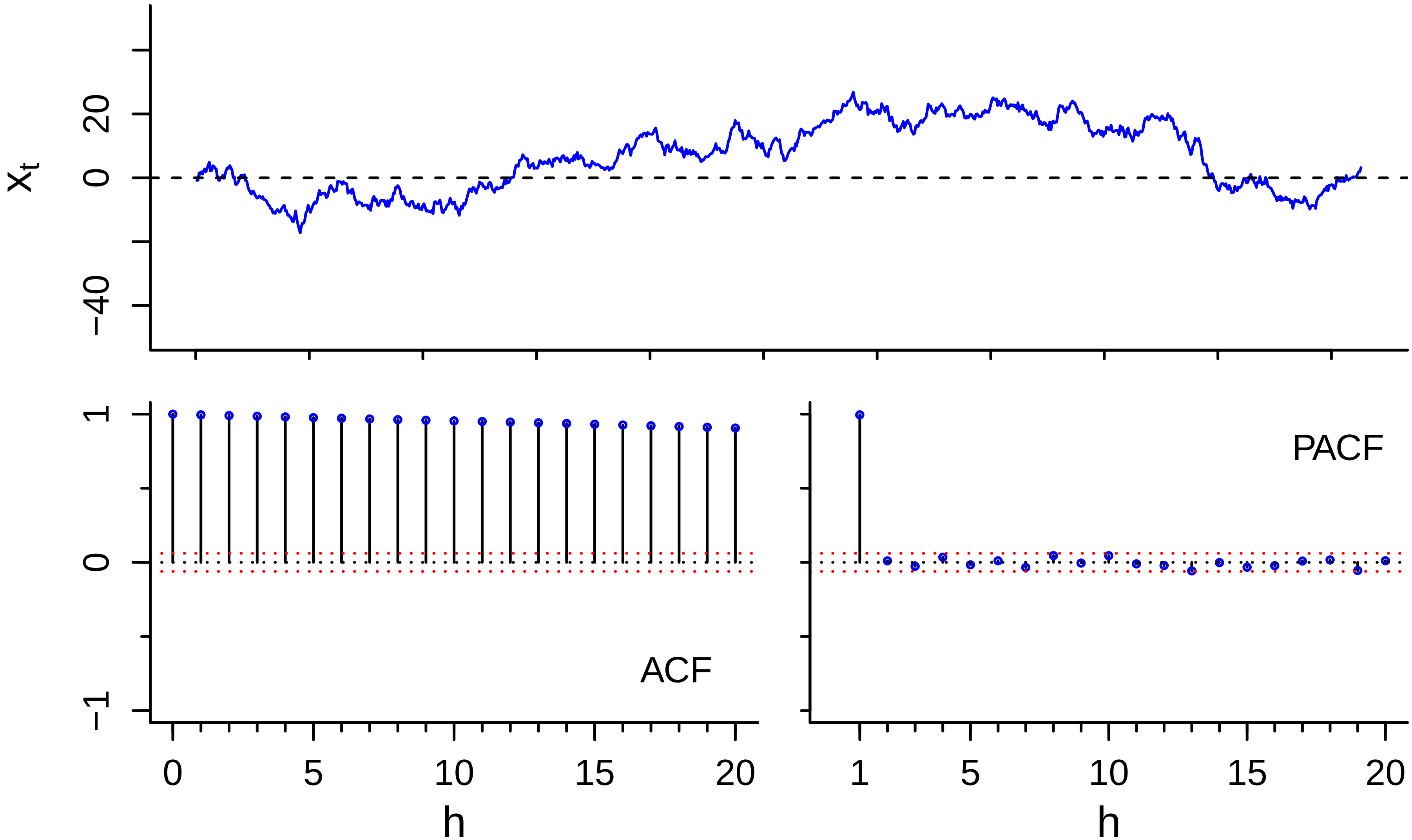
Random Walk Time Series: I



Random Walk Time Series: II



Random Walk Time Series: III



ARIMA(0,1,0) Process: III

- slowly decaying ACF is one indicator that an ARIMA model might be appropriate, but . . .
- might easily interpret sample ACF and PACF as consistent with AR(1) model with ϕ close to unity:

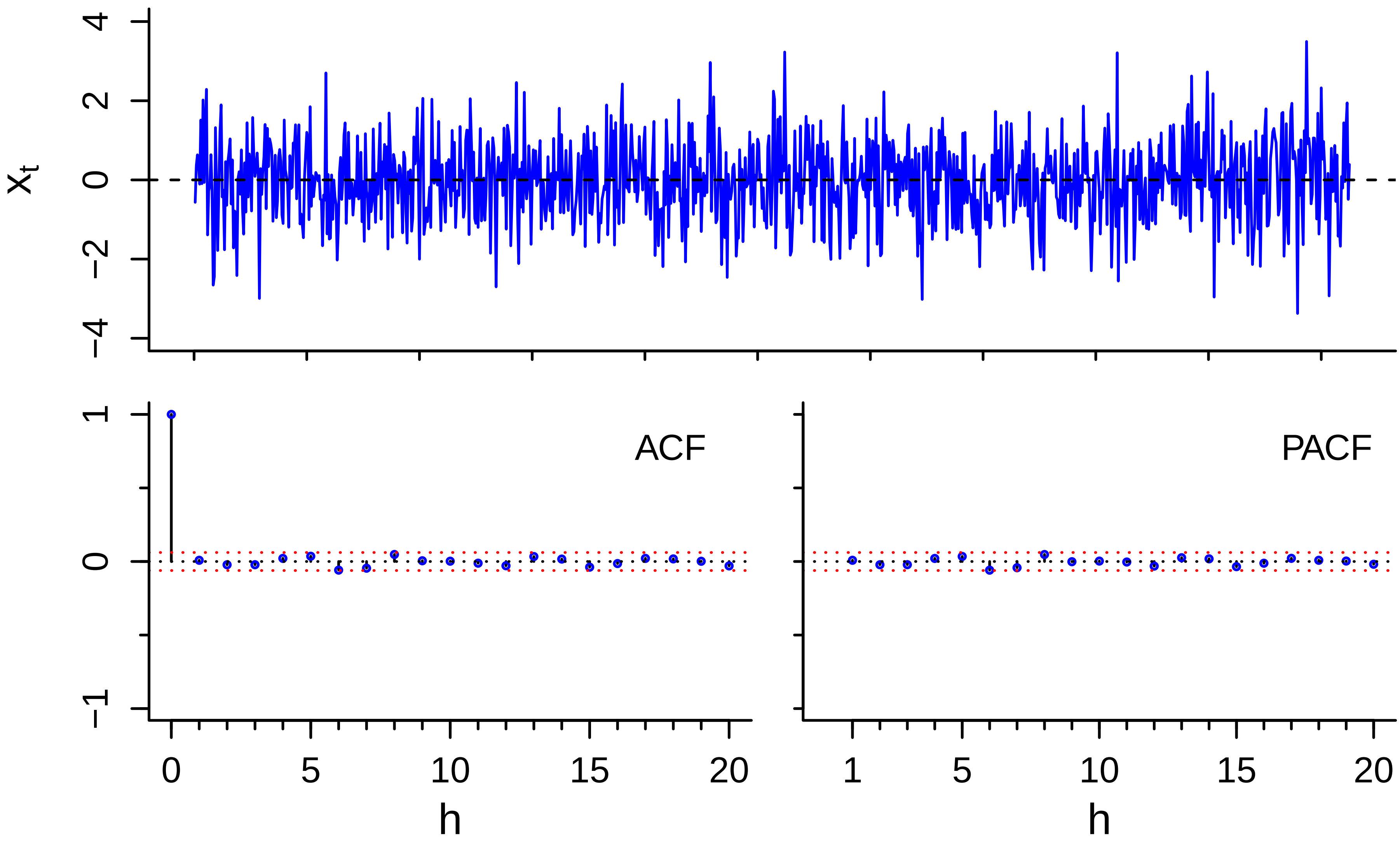
$$(1 - \phi B)X_t = X_t - \phi X_{t-1} = Z_t$$

versus

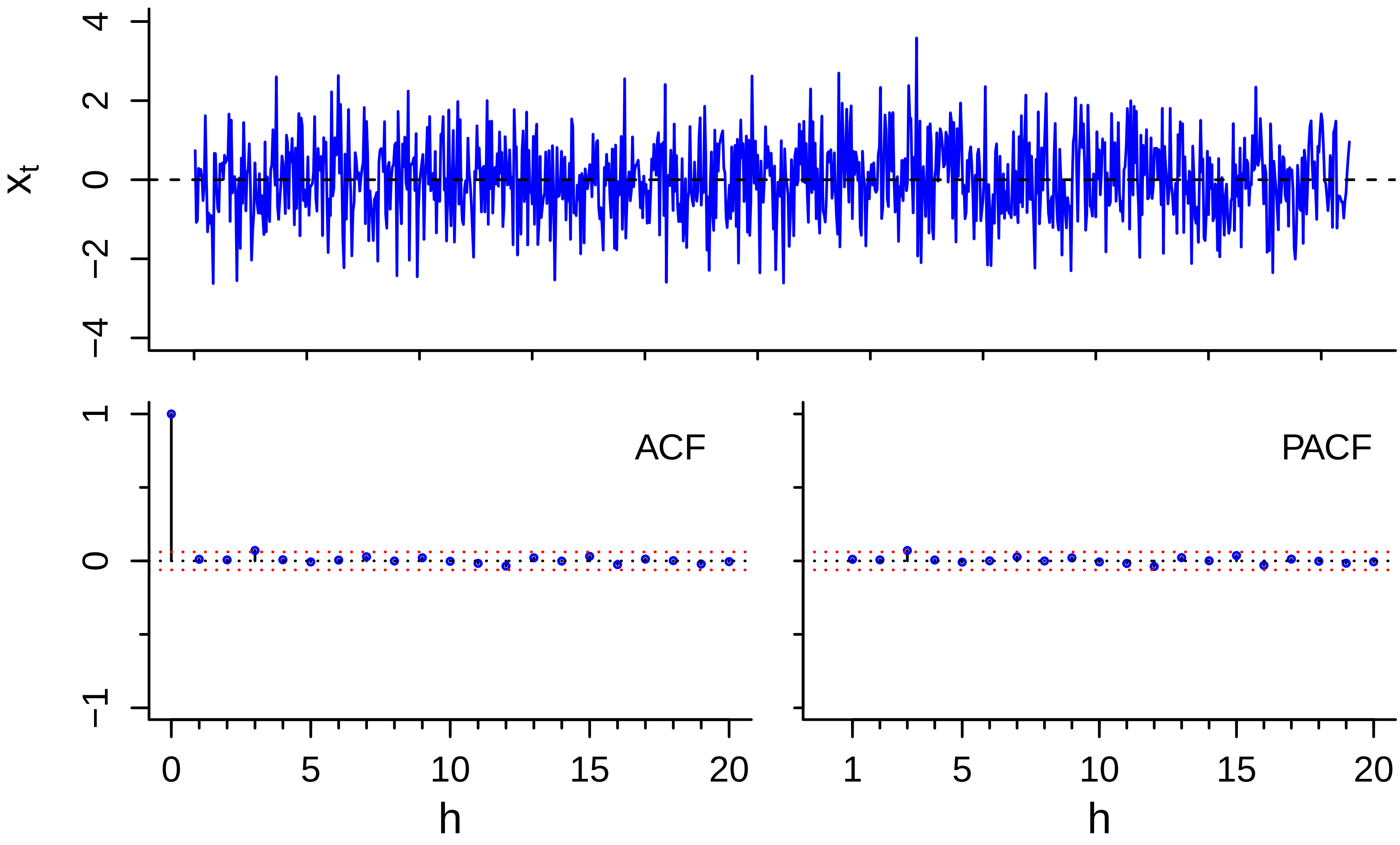
$$(1 - B)X_t = X_t - X_{t-1} = Z_t$$

- looking at ∇X_t and its sample ACF & PACF might strengthen case for ARIMA versus ARMA
- in case of random walk, ∇X_t is a white noise process

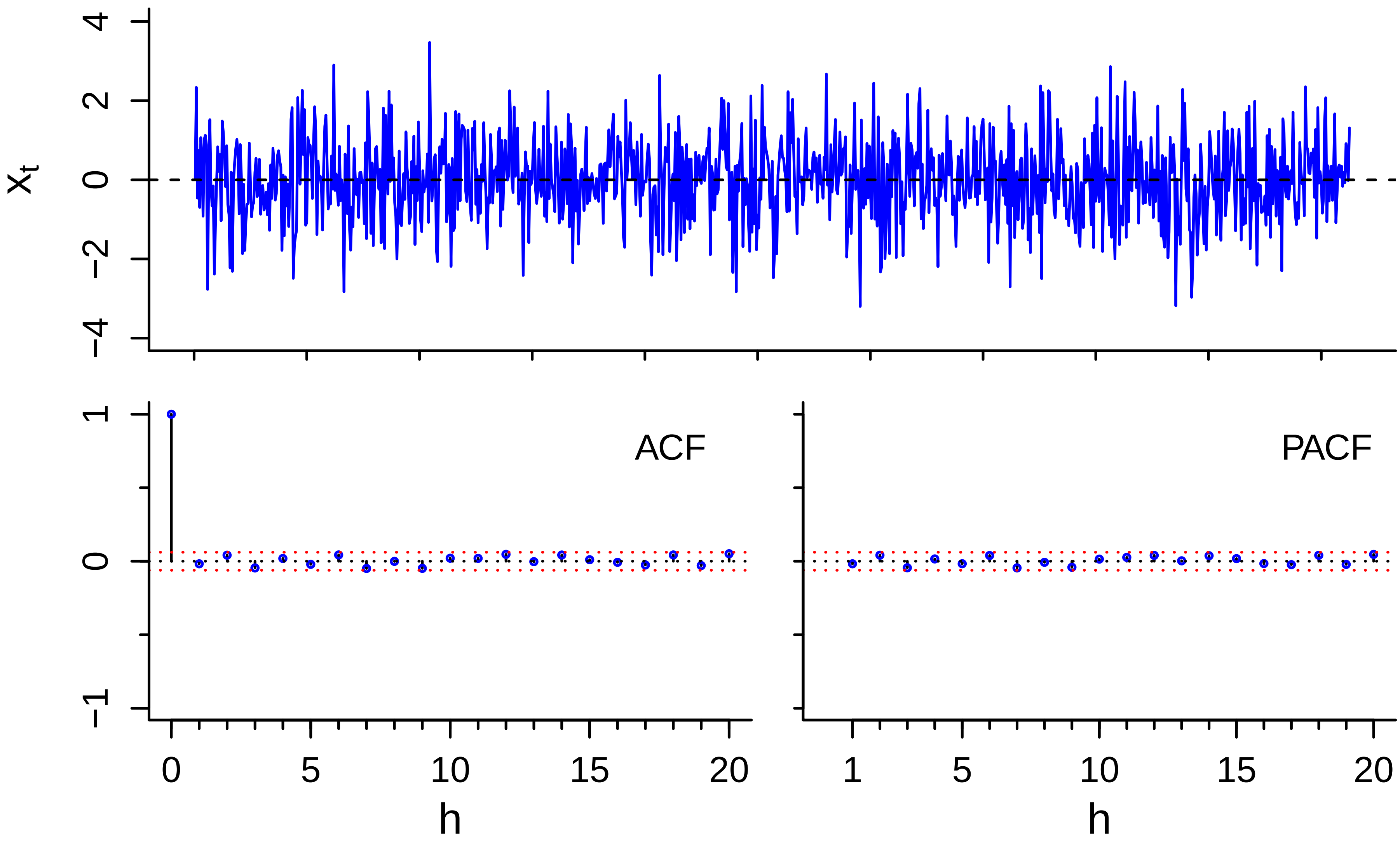
1st Difference of Random Walk Time Series: I



1st Difference of Random Walk Time Series: II



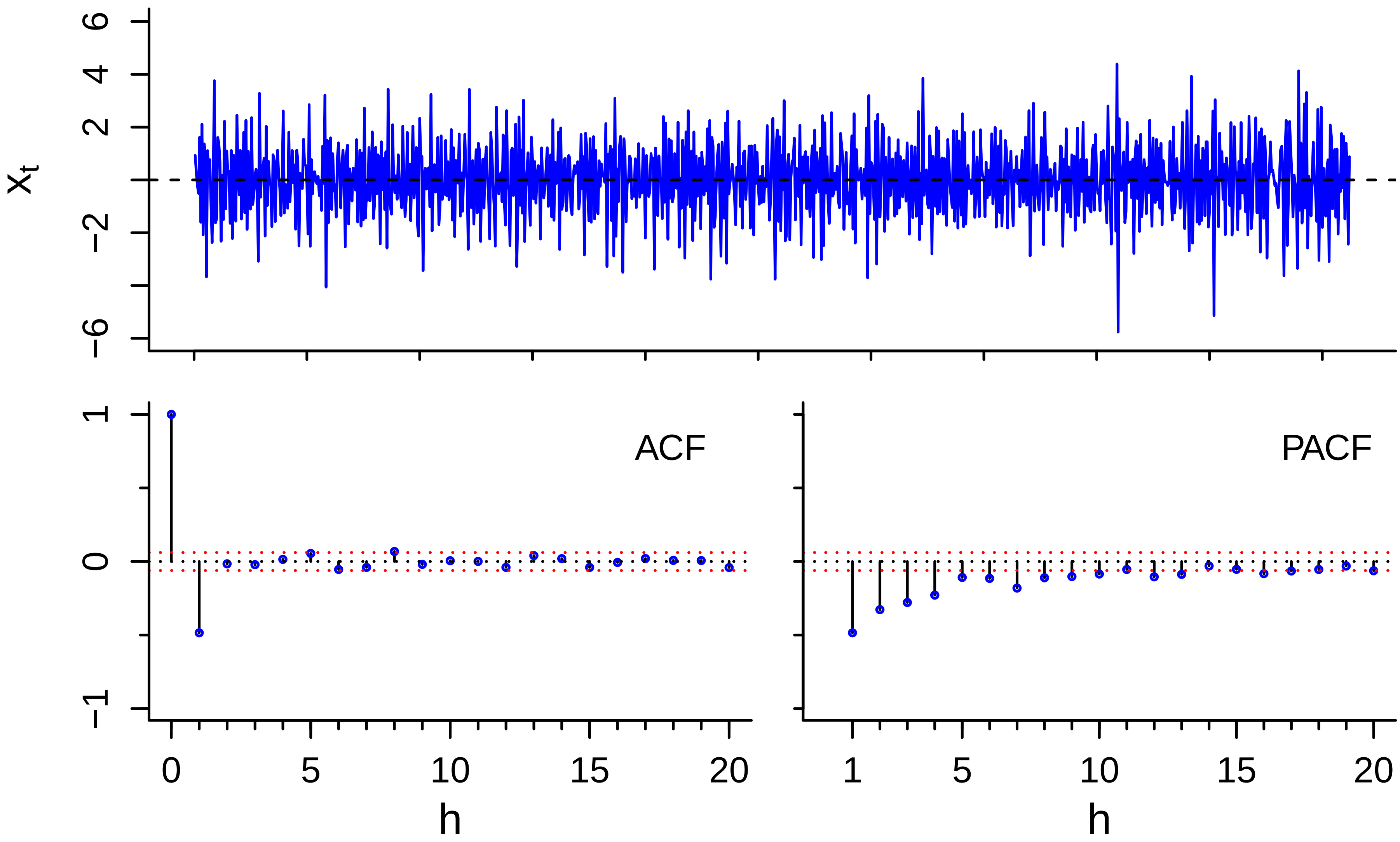
1st Difference of Random Walk Time Series: III



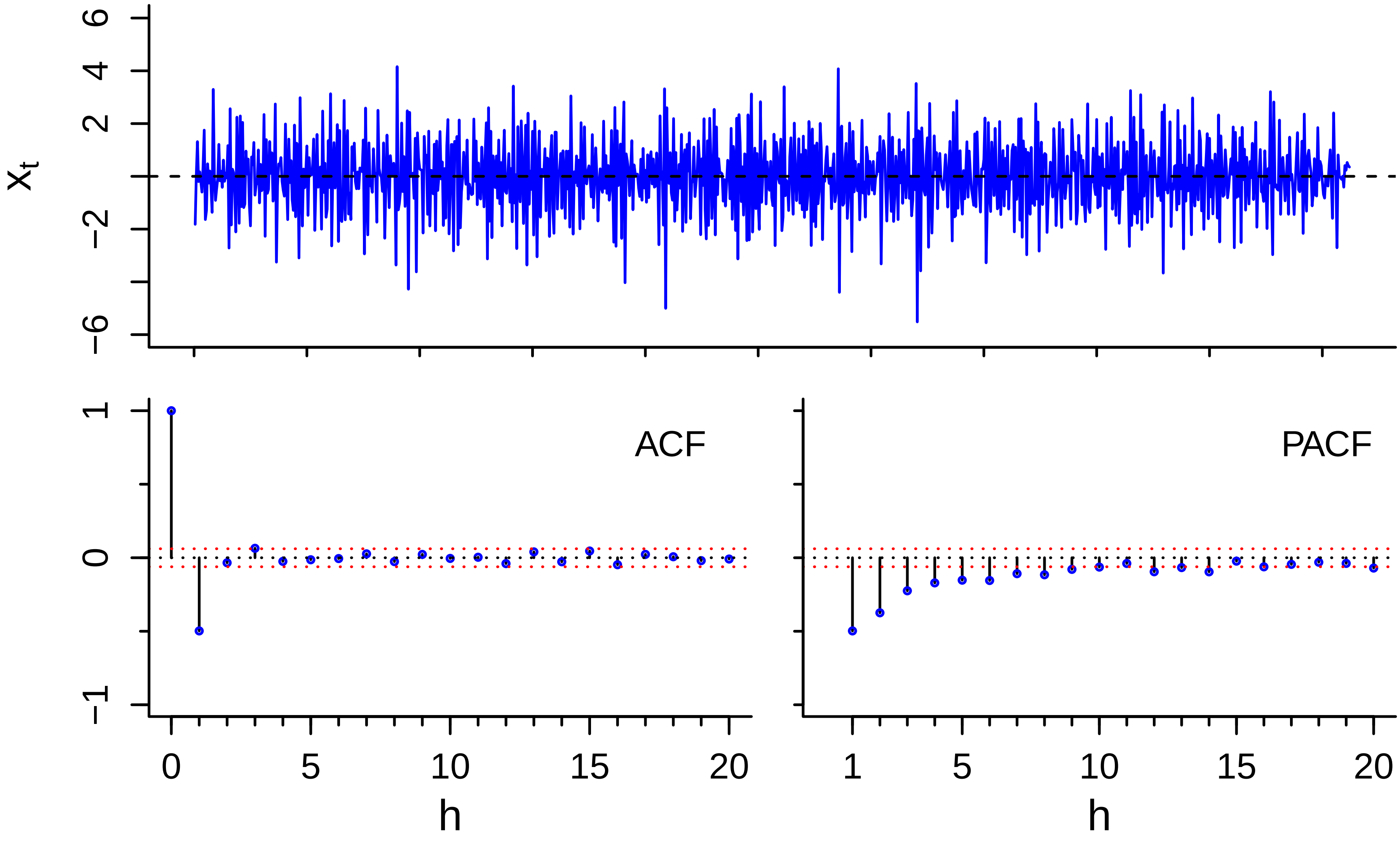
Overdifferencing: I

- while differencing a time series often seems to yield a series visually more amenable to modeling as a stationary process, overdifferencing is a danger
- 2nd difference of random walk same as 1st difference of white noise – necessarily has more complicated covariance structure (white noise takes home the prize for being process with simplest structure!)

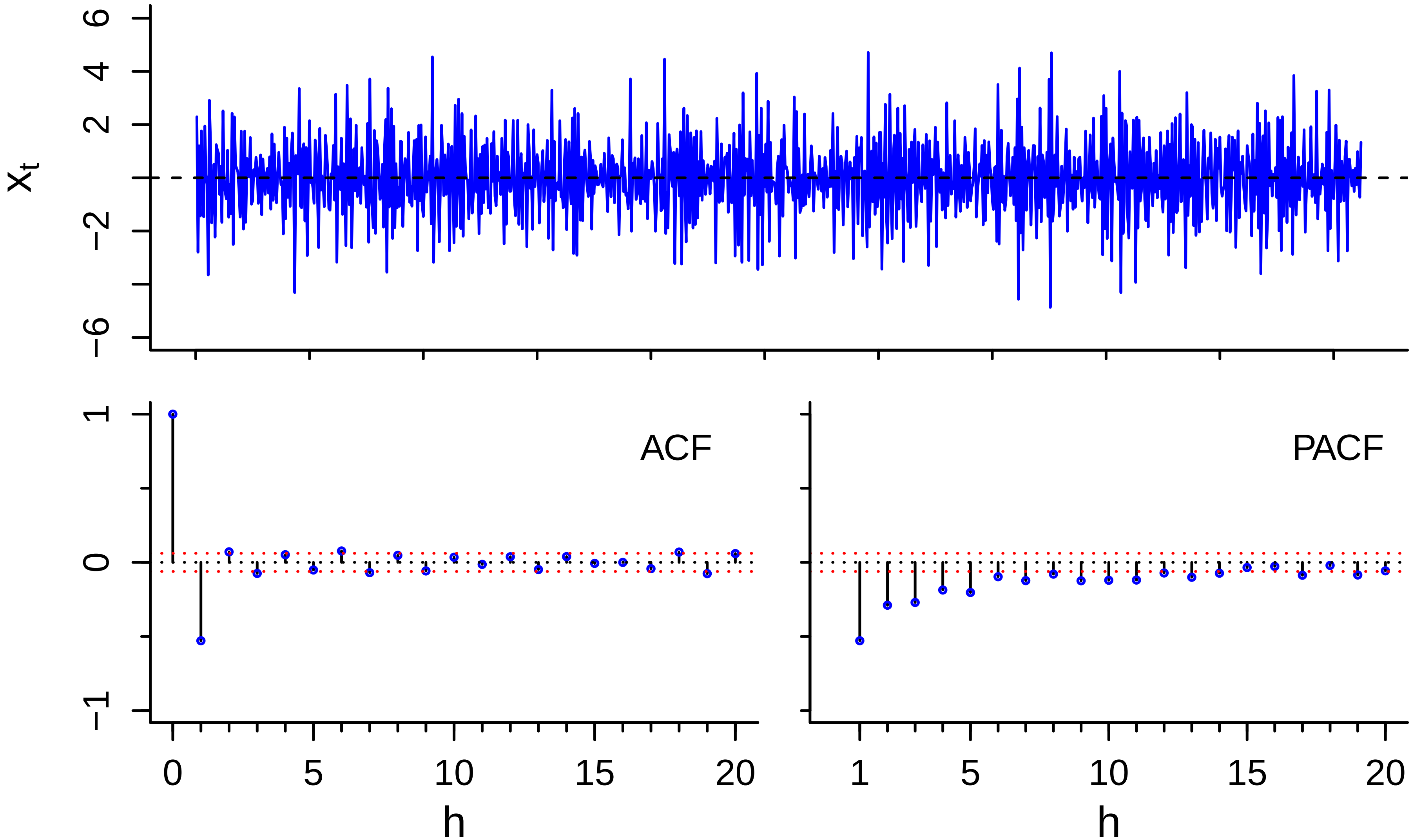
2nd Difference of Random Walk Time Series: I



2nd Difference of Random Walk Time Series: II



2nd Difference of Random Walk Time Series: III



Overdifferencing: II

- if X_t is ARMA(p, q) and hence satisfies
$$\phi(B)X_t = \theta(B)Z_t,$$
then
$$\begin{aligned}(1-B)\phi(B)X_t &= (1-B)\theta(B)Z_t\\ \phi(B)(1-B)X_t &= \theta(B)(1-B)Z_t;\end{aligned}$$
i.e., 1st difference $Y_t = \nabla X_t$ satisfies
$$\phi(B)Y_t = \theta^*(B)Z_t,$$
where $\theta^*(z) = \theta(z)(1-z)$ has a *unit root* (i.e., one of its roots is *on* the unit circle since $\theta^*(1) = 0$)

- Y_t process is thus non-invertible ARMA$(p, q+1)$, which means that π_j weights for Y_t need *not* be an absolutely summable sequence, leading to complications in 'munching' $\{Y_t\}$ into white noise (i.e., in forming one-step-ahead forecasts)

Overdifferencing: III

- evils of overdifferencing
  1. ARMA($p, q+1$) model usually has 'more complex' covariance structure than ARMA(p, q) model
  2. ARMA($p, q+1$) model has one more parameter to estimate than ARMA(p, q) model
  3. prediction of non-invertible processes tricky (but doable)
  4. estimators other than MLEs can perform quite poorly
  5. sample size reduced by one (OK, not a big deal, but . . .)

Unit Root Tests: I

- unit root tests help determine if differencing is needed
- suppose X_t obeys an ARIMA$(0, 1, 0)$ model:
$$(1 - B)X_t = X_t - X_{t-1} = Z_t$$
- above resembles the AR(1) model
$$(1 - \phi_1 B)X_t = X_t - \phi_1 X_{t-1} = Z_t$$
with $\phi(z) = (1 - \phi_1 z)$ having a unit root when $\phi_1 = 1$
- condition $\phi(1) = 0$ is equivalent to $\phi_1 = 1$
- want to devise a test for null hypothesis $\phi_1 = 1$
- obvious candidate for test statistic is $\hat{\phi}_1$, which is approximately $\mathcal{N}(\phi_1, \frac{1-\phi_1^2}{n})$ for $|\phi_1| < 1$ & large n (alas, not useful here)
- Dickey & Fuller (1979) devised an alternative test statistic

Unit Root Tests: II

- since $X_t = \phi_1 X_{t-1} + Z_t$ and $X_{t-1} = \phi_1 X_{t-2} + Z_{t-1}$, have

$$\begin{aligned}\nabla X_t &= \phi_1 \nabla X_{t-1} + \nabla Z_t \\ &= \phi_1 X_{t-1} - \phi_1 X_{t-2} + Z_t - Z_{t-1} \\ &= \phi_1 X_{t-1} - X_{t-1} + Z_t \\ &= (\phi_1 - 1) X_{t-1} + Z_t = \phi_1' X_{t-1} + Z_t, \ \text{ where } \phi_1' \stackrel{\text{def}}{=} \phi_1 - 1\end{aligned}$$

- unit root condition $\phi_1 = 1$ is equivalent to $\phi_1' = 0$
- Dickey–Fuller unit root test: use ordinary least squares (OLS) to regress ∇X_t on X_{t-1}, $t = 2, 3, \ldots, n$, and then test null hypothesis $\phi_1' = 0$
- for AR(1) model with $E\{X_t\} = \mu \neq 0$, model becomes

$$\nabla X_t = \phi_0' + \phi_1' X_{t-1} + Z_t,$$

where $\phi_0' = \mu(1 - \phi_1)$ and, as before, $\phi_1' = \phi_1 - 1$

Unit Root Tests: III

- in regression model $y_t = a + bx_t + e_t$, OLS estimator of b is

$$\hat{b} = \frac{\sum_t (x_t - \bar{x})(y_t - \bar{y})}{\sum_t (x_t - \bar{x})^2} = \frac{\sum_t (x_t - \bar{x}) y_t}{\sum_t (x_t - \bar{x})^2},$$

  where $\bar{x}$ & $\bar{y}$ are sample means

- given $X_1, \ldots, X_n$ & letting $\widetilde{X}_{t-1} \stackrel{\text{def}}{=} X_{t-1} - \frac{1}{n-1} \sum_{s=1}^{n-1} X_s$, let $\hat{\phi}_1'$ be OLS estimator of ϕ_1' in model $\nabla X_t = \phi_0' + \phi_1' X_{t-1} + Z_t$:

$$\hat{\phi}_1' = \frac{\sum_{t=2}^n \widetilde{X}_{t-1} \nabla X_t}{\sum_{t=2}^n \widetilde{X}_{t-1}^2} = \frac{\sum_{t=2}^n \widetilde{X}_{t-1}(X_t - X_{t-1})}{\sum_{t=2}^n \widetilde{X}_{t-1}^2}$$

$$= \frac{\sum_{t=2}^n \widetilde{X}_{t-1}(\widetilde{X}_t - \widetilde{X}_{t-1})}{\sum_{t=2}^n \widetilde{X}_{t-1}^2} = \frac{\sum_{t=2}^n \widetilde{X}_{t-1}\widetilde{X}_t}{\sum_{t=2}^n \widetilde{X}_{t-1}^2} - 1$$

Unit Root Tests: IV

- note connection of

$$\hat{\phi}_1' = \frac{\sum_{t=2}^{n} \widetilde{X}_{t-1}\widetilde{X}_t}{\sum_{t=2}^{n} \widetilde{X}_t^2} - 1$$

$$= \frac{\sum_{t=2}^{n}(X_{t-1} - \frac{1}{n-1}\sum_{s=1}^{n-1} X_s)(X_t - \frac{1}{n-1}\sum_{s=1}^{n-1} X_s)}{\sum_{t=2}^{n}(X_t - \frac{1}{n-1}\sum_{s=1}^{n-1} X_s)^2} - 1$$

to Y–W estimator $\hat{\phi}_1$ of ϕ using $X_1, \ldots, X_n$ after centering:

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n}(X_{t-1} - \overline{X})(X_t - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

- hence $\hat{\phi}_1' \approx \hat{\phi}_1 - 1$

Unit Root Tests: V

- in regression model $y_t = a + bx_t + e_t$ for m observations, usual standard error of OLS estimator $\hat{b}$ is taken to be

$$\widehat{\mathrm{SE}}(\hat{b}) = \left(\frac{\sum_t (y_t - \hat{a} - \hat{b}x_t)^2}{(m-2)\sum_t (x_t - \bar{x})^2}\right)^{1/2}, \quad \text{where } \hat{a} = \bar{y} - \hat{b}\bar{x},$$

- for model $\nabla X_t = \phi_0' + \phi_1' X_{t-1} + Z_t$, above leads to

$$\widehat{\mathrm{SE}}(\hat{\phi}_1') = \left(\frac{\sum_{t=2}^{n} (\nabla X_t - \hat{\phi}_0' - \hat{\phi}_1' X_{t-1})^2}{(n-3)\sum_{t=2}^{n}\left(X_{t-1} - \frac{1}{n-1}\sum_{s=1}^{n-1} X_s\right)^2}\right)^{1/2}$$

where

$$\hat{\phi}_0' = \frac{1}{n-1}\left(\sum_{t=2}^{n} \nabla X_t - \hat{\phi}_1' \sum_{t=1}^{n-1} X_t\right) = \frac{1}{n-1}\left(X_n - X_1 - \hat{\phi}_1' \sum_{t=1}^{n-1} X_t\right)$$

Unit Root Tests: VI

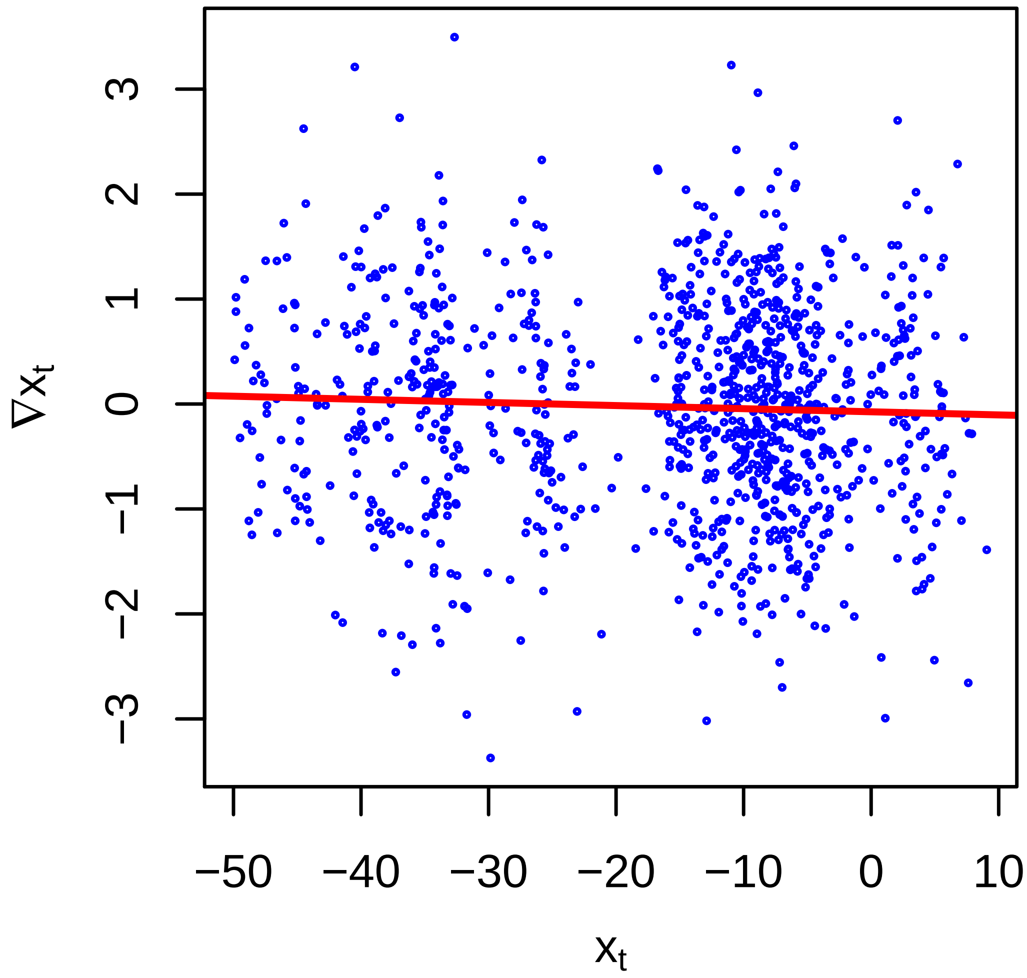
- test statistic is t-like ratio
$$t = \frac{\hat{\phi}_1'}{\widehat{\mathrm{SE}}(\hat{\phi}_1')},$$
where denominator is standard error for slope as prescribed by OLS theory (but: t does *not* obey a t-distribution!!!)
- null hypothesis is $\phi_1' = \phi_1 - 1 = 0$ (there is a unit root)
- $\phi_1' > 0$ (i.e., $\phi_1 > 1$) problematic (non-causal AR(1) model), so will take alternative to be $\phi_1' < 0$
- reject null $\phi_1' = 0$ (have a unit root) in favor of alterative $\phi_1' < 0$ (AR(1) is appropriate) at level, say, $\alpha = 0.05$ if t falls below 5% percentage point established for Dickey–Fuller test statistic under assumption that n is large

Unit Root Tests: VII

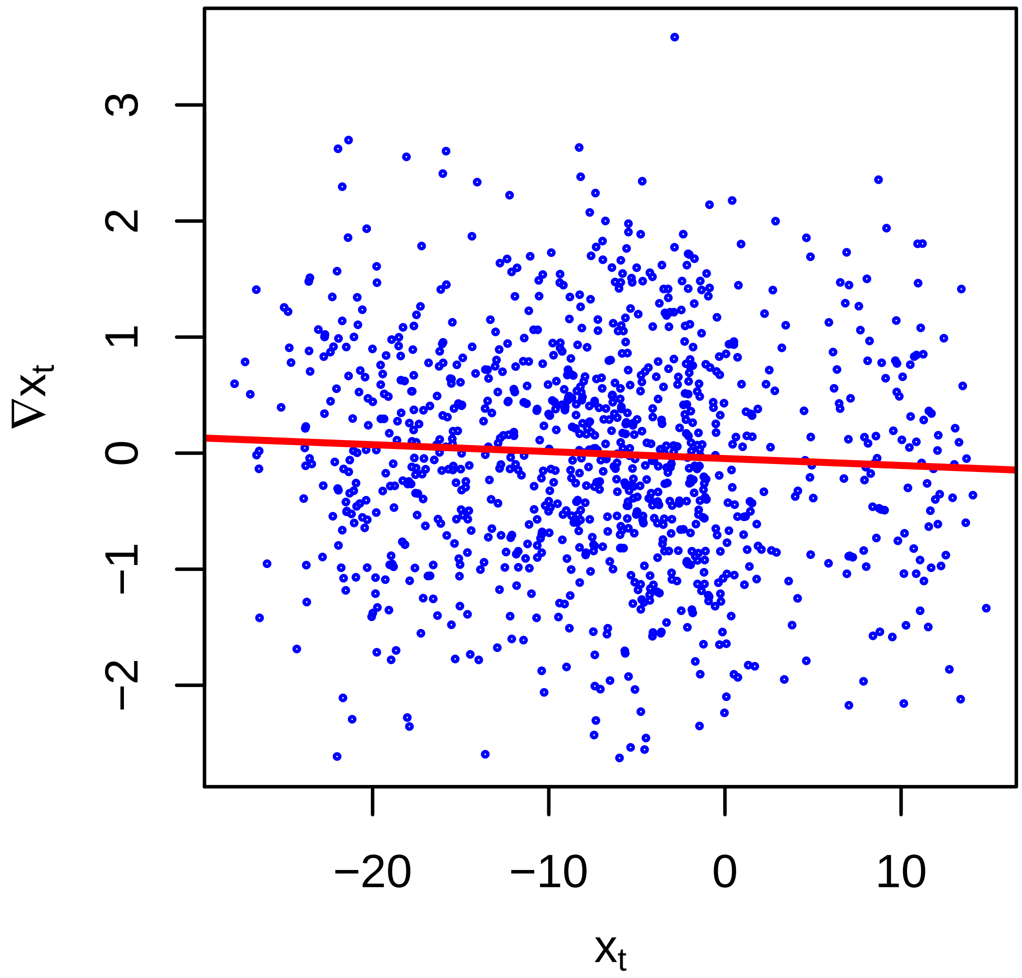
- let's demonstrate test by lobbing it some softballs: 3 random walk series (overheads XIV–6 to XIV–8) and 3 white noise series (overheads XIV–10 to XIV–12)
- 1%, 5% and 10% percentage points are -3.43, -2.86 and -2.57

	$\hat{\phi}_1'$	$\widehat{\mathrm{SE}}(\hat{\phi}_1')$	t	null hypothesis
random walk #1	-0.002963	0.002185	-1.357	fail to reject
random walk #2	-0.005964	0.003353	-1.779	fail to reject
random walk #3	-0.004671	0.002961	-1.577	fail to reject
white noise #1	-0.99170	0.03128	-31.71	reject
white noise #2	-0.98890	0.03129	-31.61	reject
white noise #3	-1.01723	0.03130	-32.50	reject

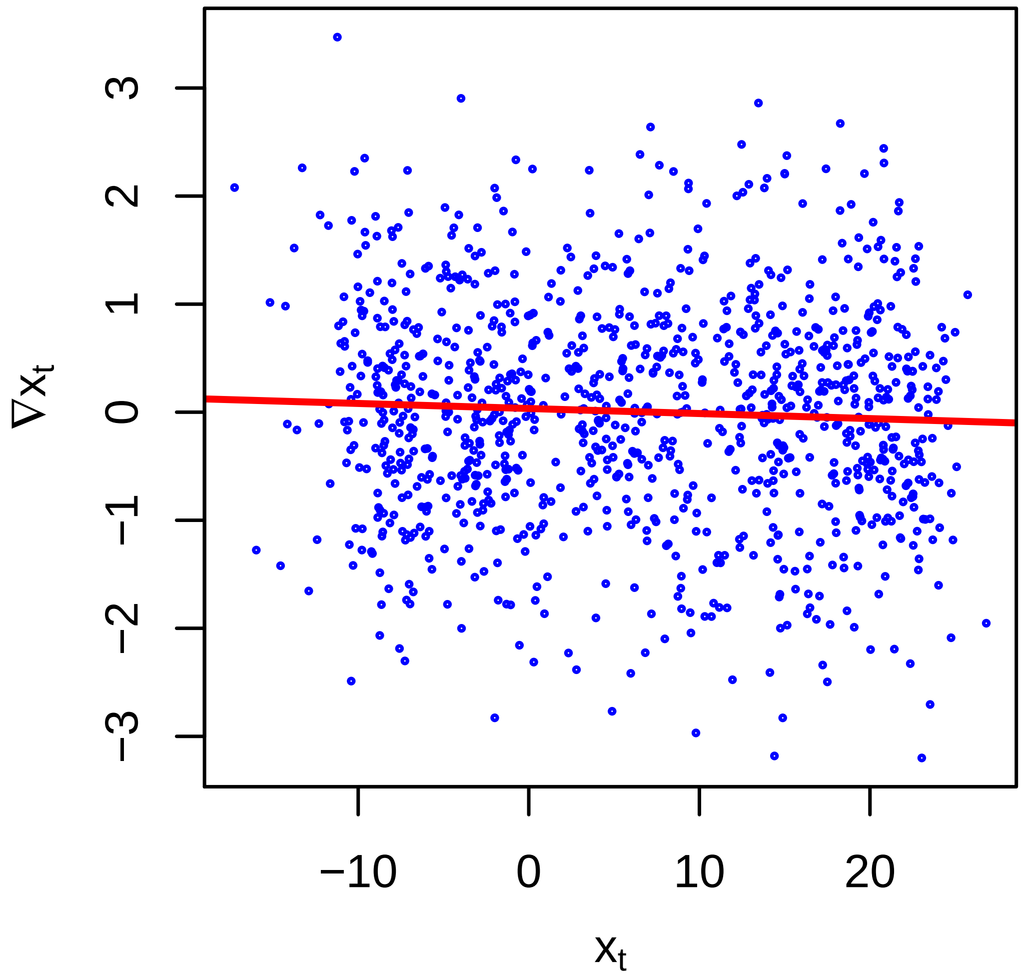
Scatterplot and Fitted Line for Random Walk: I



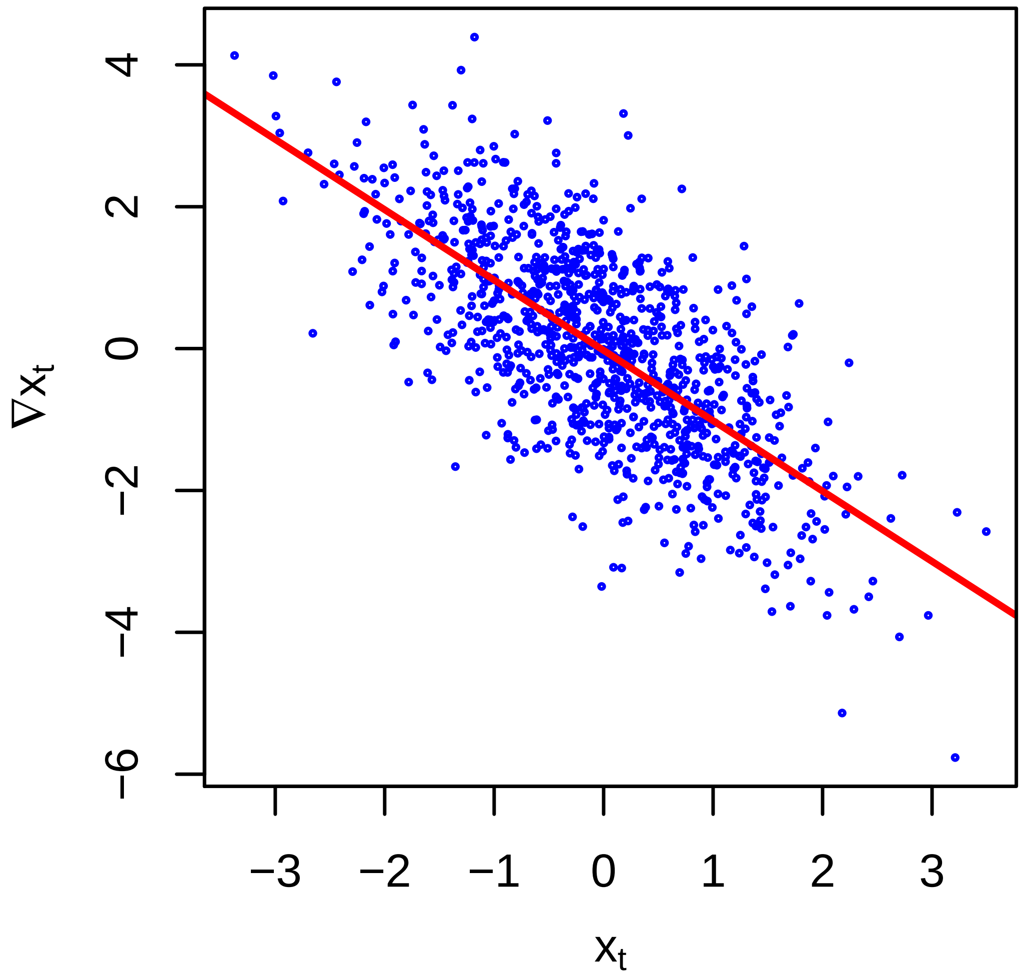
Scatterplot and Fitted Line for Random Walk: II



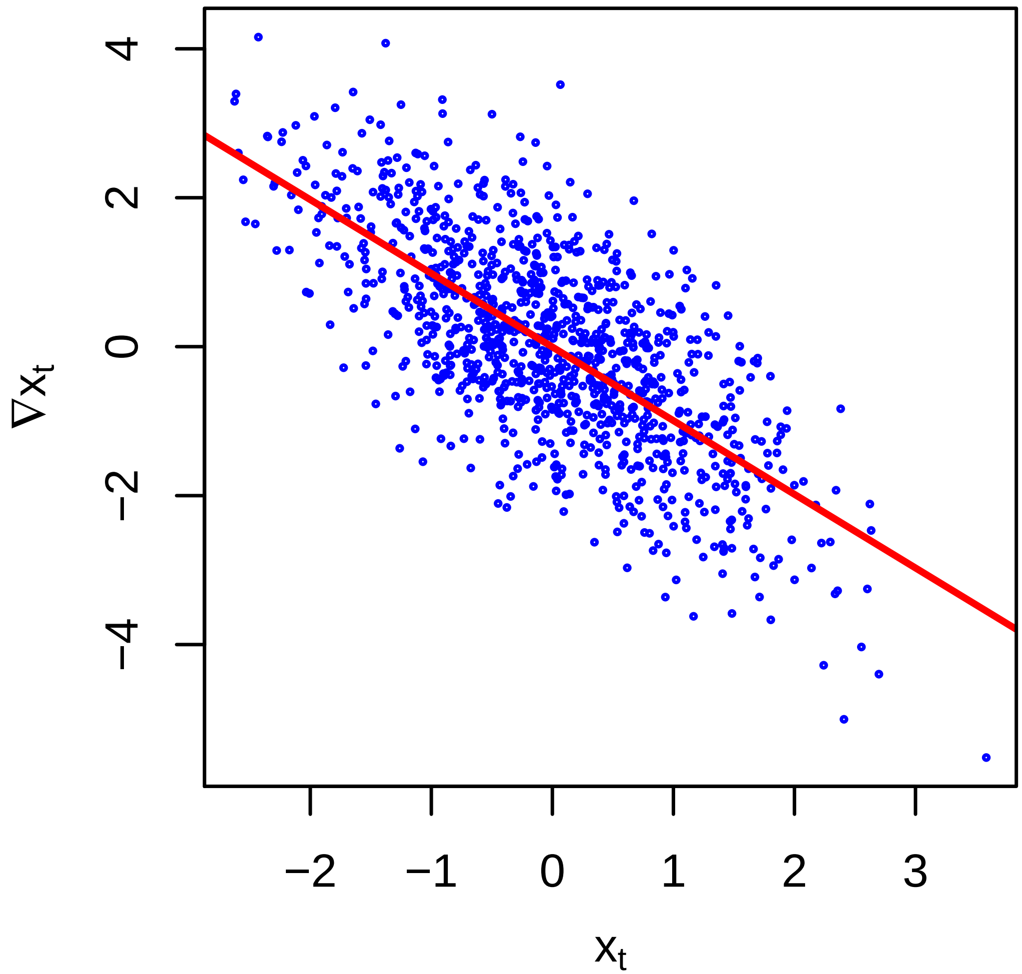
Scatterplot and Fitted Line for Random Walk: III



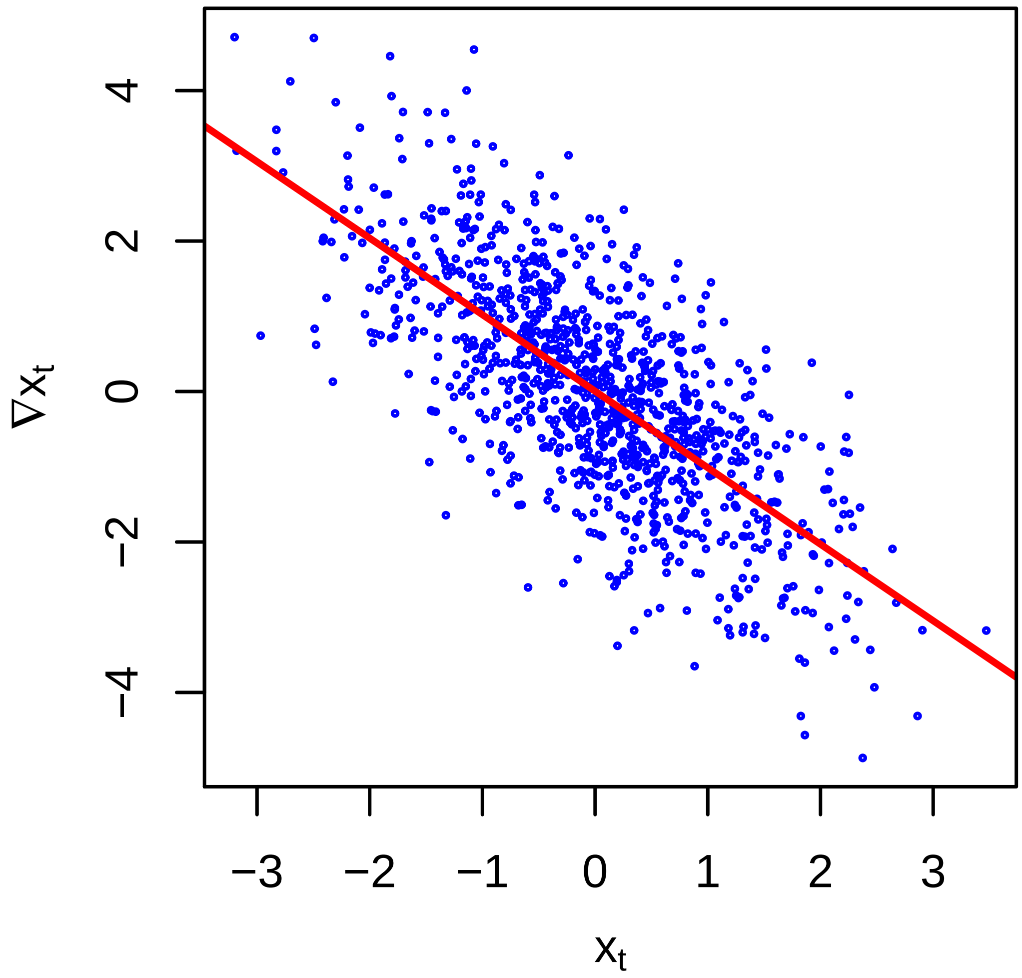
Scatterplot and Fitted Line for White Noise: I



Scatterplot and Fitted Line for White Noise: II



Scatterplot and Fitted Line for White Noise: III



Unit Root Tests: VIII

- idea for testing ARIMA(0,1,0) vs. AR(1) (i.e., ARIMA(1,0,0)) can be extended to test ARIMA($p-1$,1,0) vs. AR(p)
- regard ARIMA($p-1, 1, 0$) model
$$\phi^*(B)(1-B)X_t = Z_t,$$
where
$$\phi^*(z) = 1 - \phi_1^* z - \cdots - \phi_{p-1}^* z^{p-1},$$
as limiting case of AR(p) model
$$\phi(B)X_t = Z_t,$$
where
$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$
by equating $\phi(z)$ with $\phi^*(z)(1-z)$
- unit root condition $\phi(1) = 0$ now equivalent to $1 - \sum_{i=1}^{p} \phi_i = 0$

Unit Root Tests: IX

- cleverness (Dickey & Fuller, 1979) says

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

can be rewritten as

$$\nabla X_t = \phi_1' X_{t-1} + \phi_2' \nabla X_{t-1} + \cdots + \phi_p' \nabla X_{t-(p-1)} + Z_t, \quad (*)$$

where $\phi_1' = \sum_{i=1}^{p} \phi_i - 1$ and $\phi_j' = -\sum_{i=j}^{p} \phi_i - 1$, $j = 2, \ldots, p$

- adding intercept term ϕ_0' to $(*)$ to allow for AR(p) model with nonzero mean yields multiple regression model

$$\nabla X_t = \phi_0' + \phi_1' X_{t-1} + \phi_2' \nabla X_{t-1} + \cdots + \phi_p' \nabla X_{t-(p-1)} + Z_t,$$

where $\phi_1' = 0$ is equivalent to unit root condition

Unit Root Tests: X

- as before, use OLS to get estimator $\hat{\phi}_1'$ and form t-like ratio to test null hypothesis $\phi_1' = 0$
- above called *augmented* Dickey–Fuller (ADF) unit root test
- note: unit root tests have also been devised for $\theta(z)$ to assess possible noninvertibility
- Chapter 4 of Zivot & Wang (2006) has additional discussion

Nonstationarity Due to Polynomial Trend: I

- consider time series given by
$$Y_t = a + bt + X_t \text{ with } b \neq 0,$$
where X_t is ARMA(p,q) with zero mean: $\phi(B)X_t = \theta(B)Z_t$
- Y_t is nonstationary since $E\{Y_t\} = a + bt$
- first difference of Y_t is stationary:
$$\nabla Y_t = Y_t - Y_{t-1} = a + bt + X_t - (a + b(t-1) + X_{t-1}) = b + \nabla X_t,$$
where ∇X_t is necessarily a stationary process
- since $(1 - B)\phi(B)X_t = (1 - B)\theta(B)Z_t$, can be reexpressed as $\phi(B)\nabla X_t = (1 - B)\theta(B)Z_t$, it follows that ∇X_t is a noninvertible ARMA(p,$q+1$) process
- Y_t is thus an ARIMA(p,1,$q+1$) process whose first difference is a noninvertible ARMA(p,$q+1$) process with mean b

Nonstationarity Due to Polynomial Trend: II

- when X_t is ARMA(p,q), treating

$$Y_t = a + bt + X_t$$

as an ARIMA(p,1,$q+1$) process thus leads to evils of overdifferencing (see overhead XIV–18)

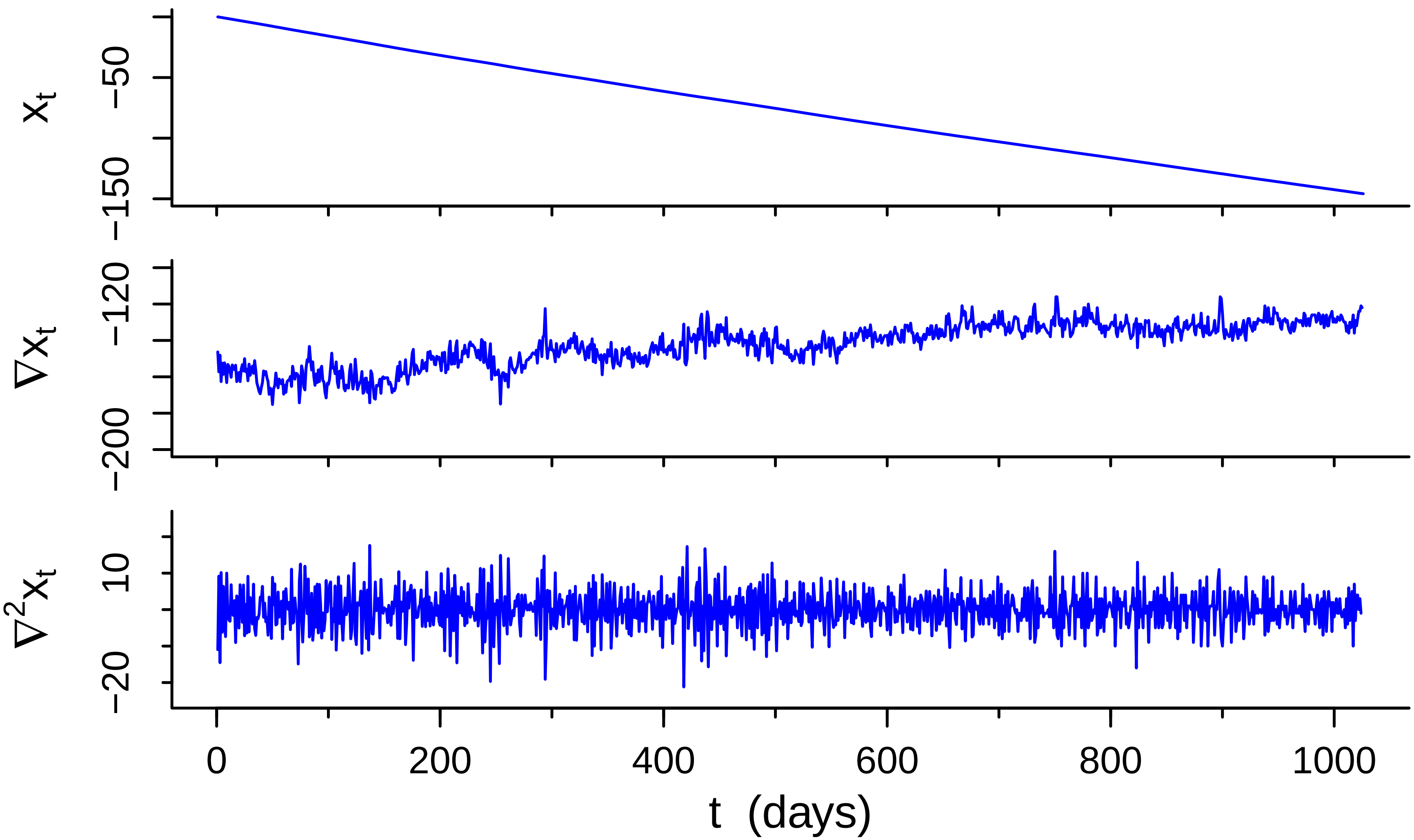
- suppose, however, that time series satisfies

$$Y_t = a + bt + X_t \text{ with } b \neq 0,$$

where X_t is an ARIMA(p,1,q) process whose first difference is an invertible ARMA(p,q) process with zero mean

- now Y_t becomes an ARIMA(p,1,q) whose first difference is an invertible ARMA(p,q) process with mean b

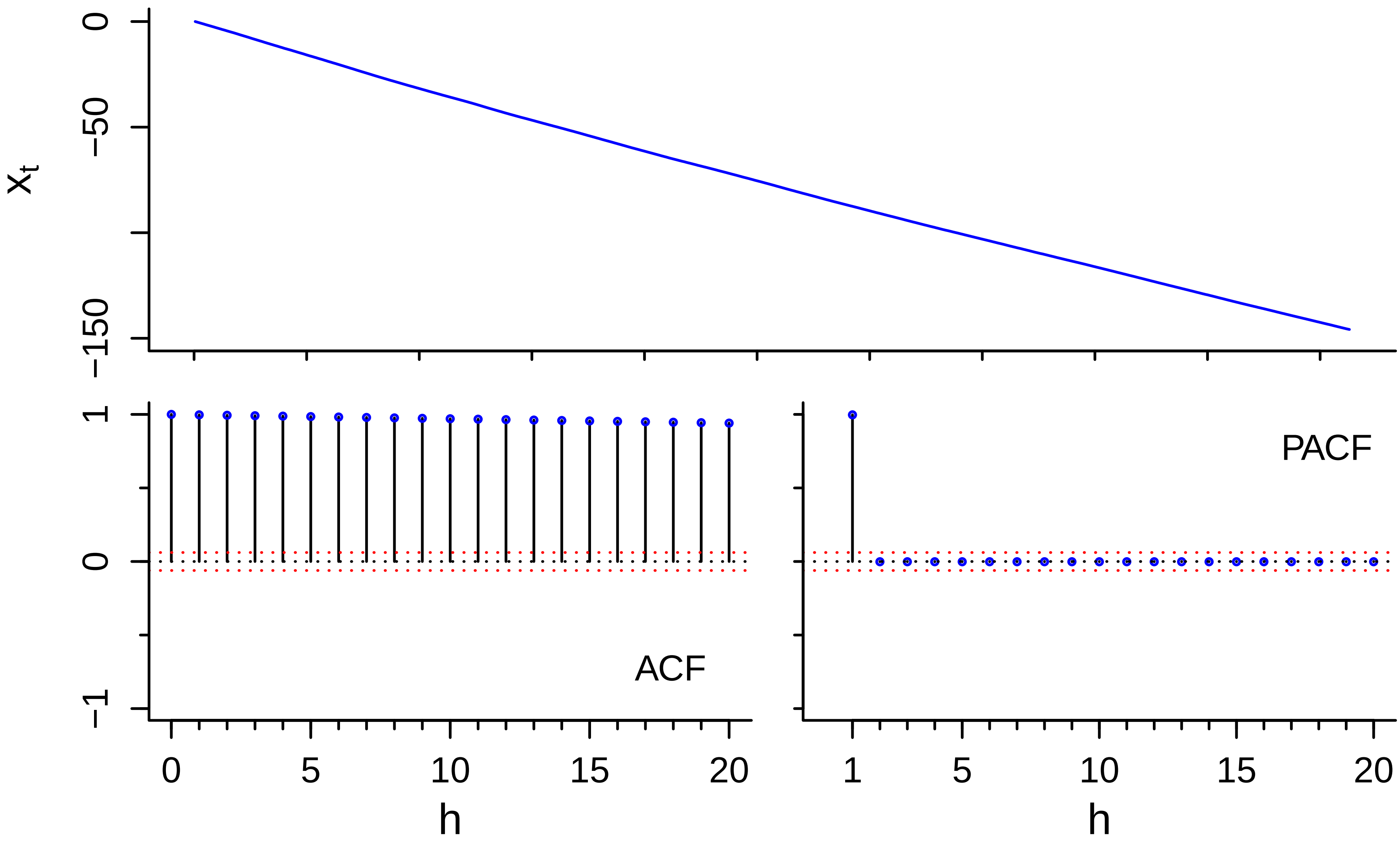
Atomic Clock Deviates: I



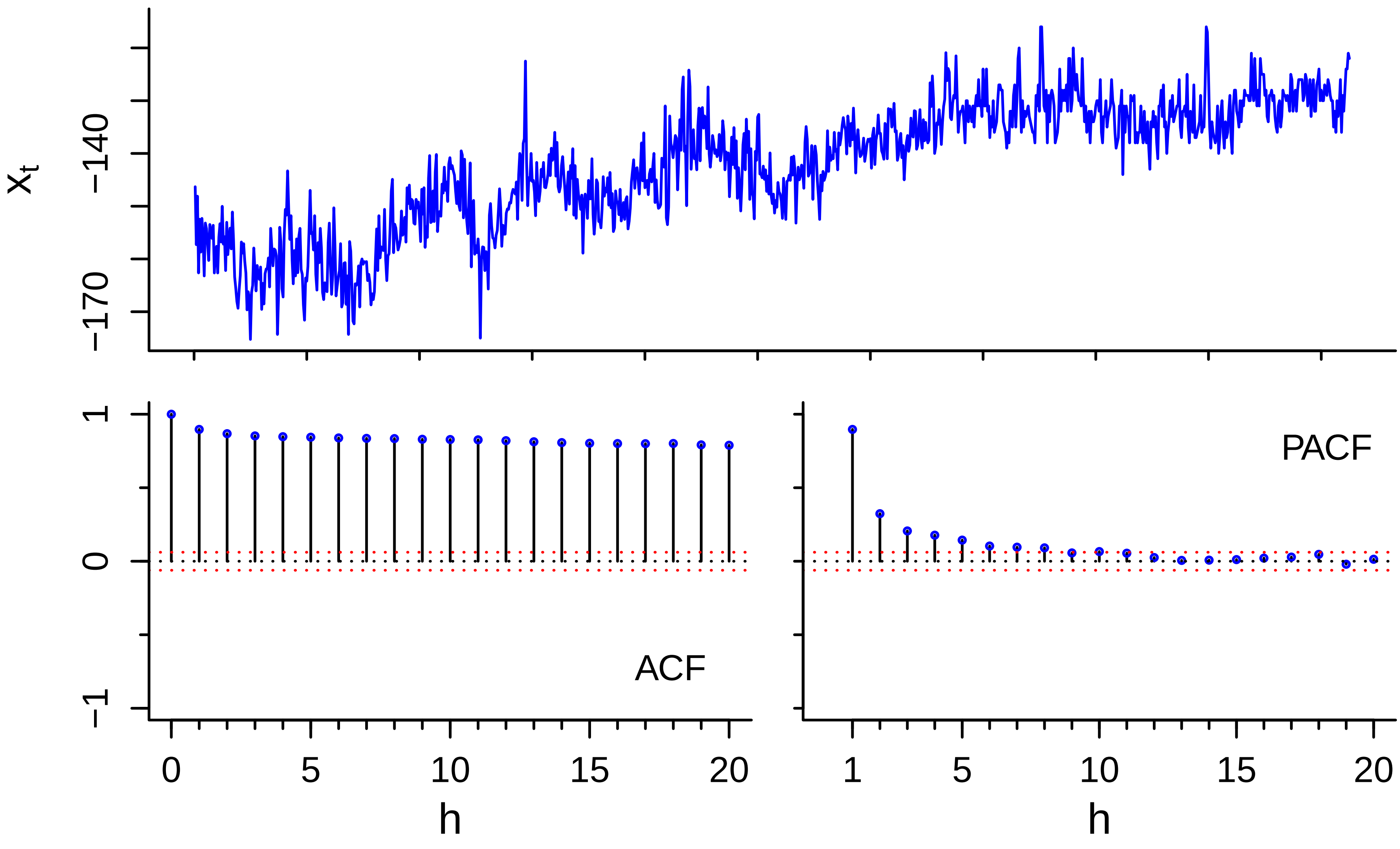
Atomic Clock Deviates: II

- top plot: errors X_t in time kept by atomic clock 571 as compared to time kept at Naval Observatory (measured in microseconds, where 1,000,000 microseconds = 1 second)
- middle: first backward differences ∇X_t in nanoseconds (1000 nanoseconds = 1 microsecond) – can be related to frequency mechanism driving clock
- bottom: second backward differences $\nabla^2 X_t$, also in nanoseconds – can be related to changes in frequency
- note: possible linear trend in ∇X_t would result in nonzero mean for $\nabla^2 X_t$,

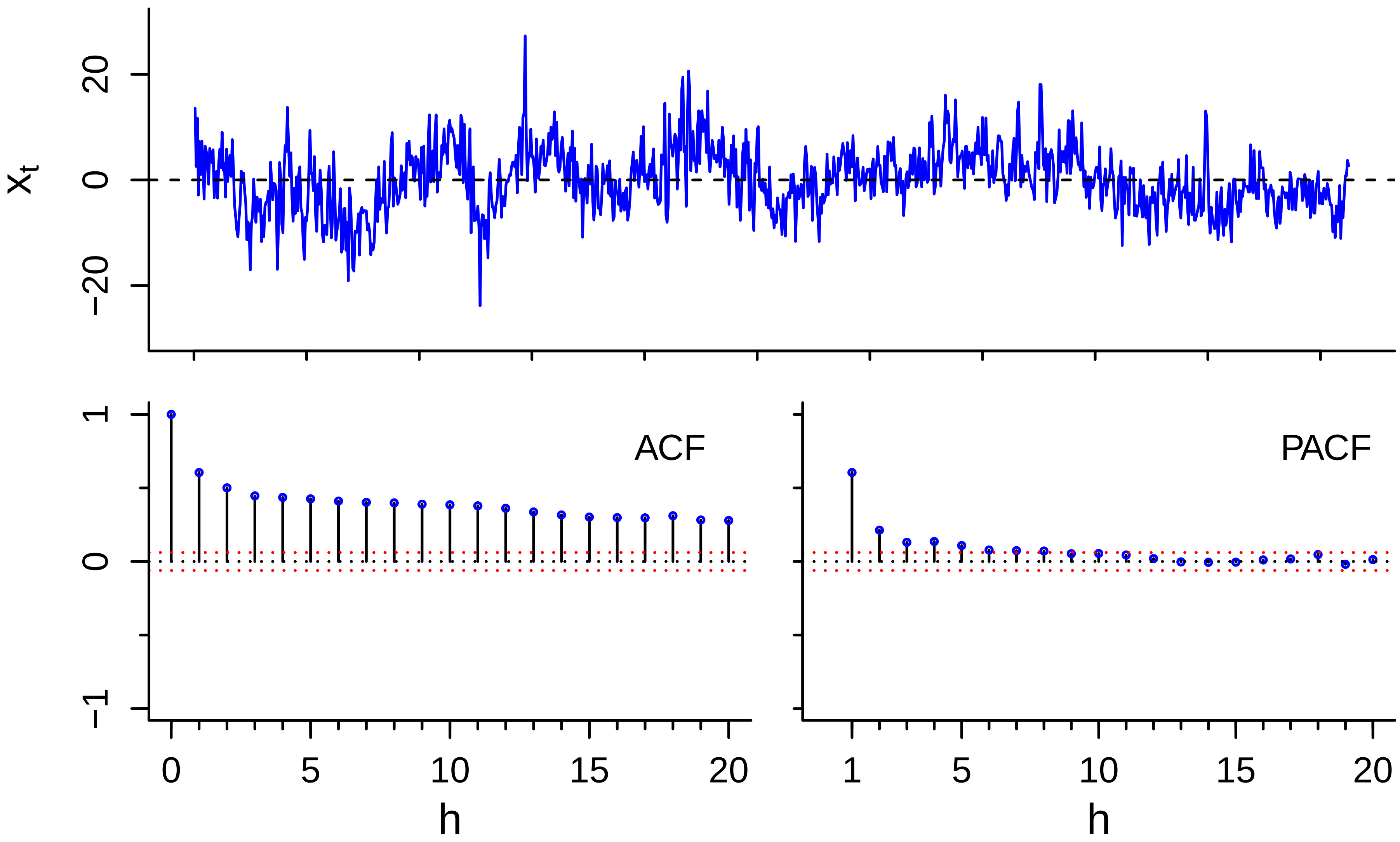
Atomic Clock Time Series



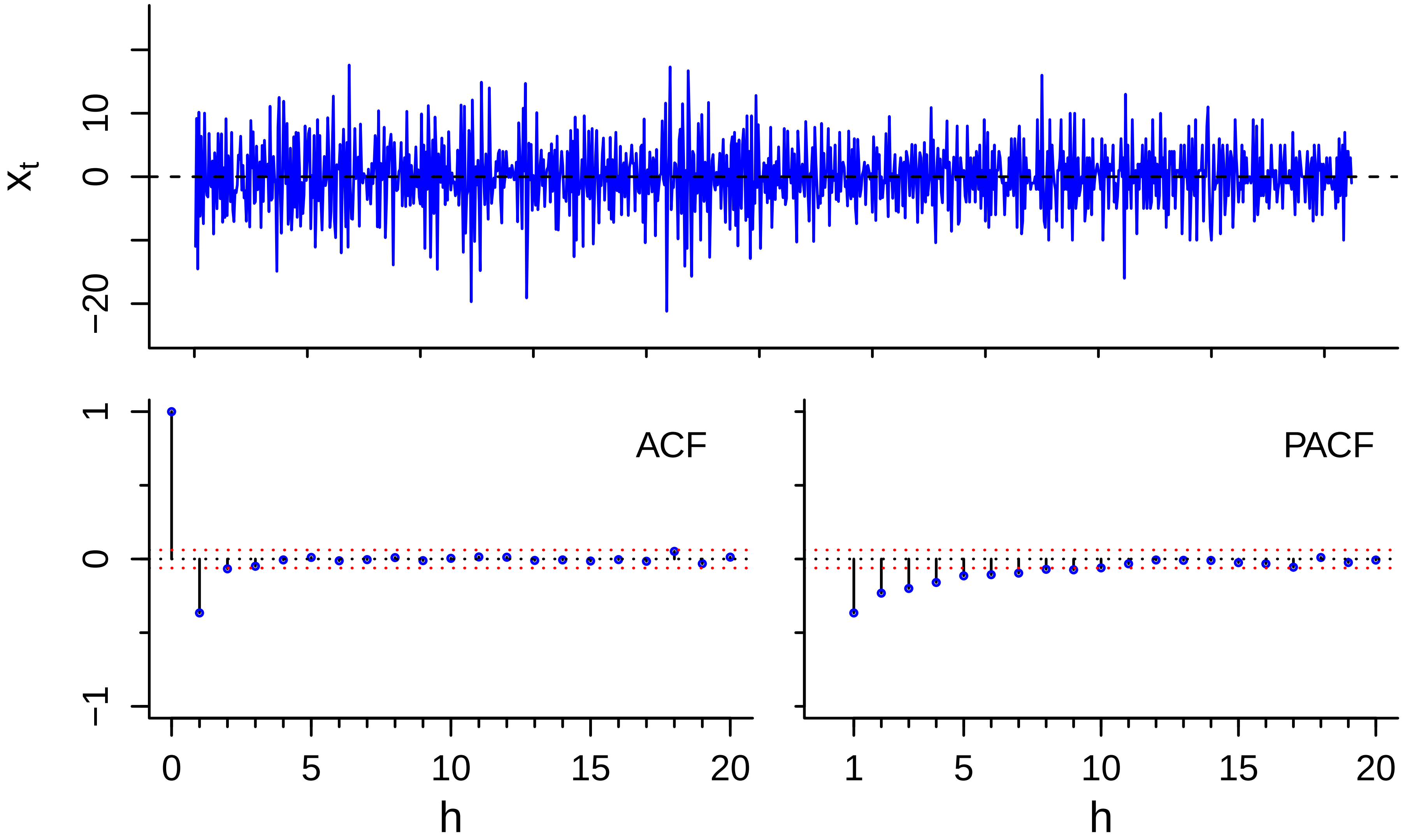
1st Difference of Atomic Clock Time Series



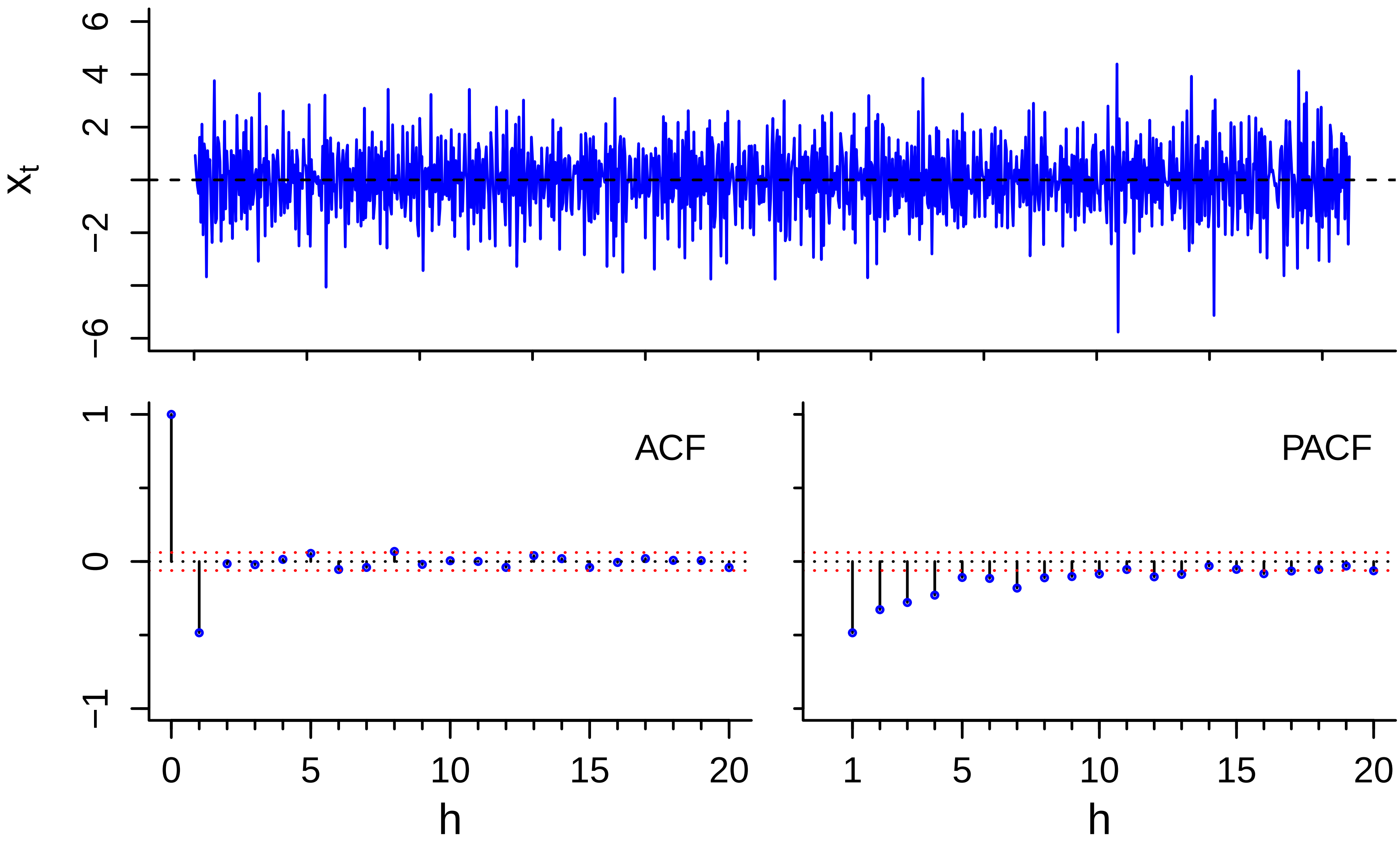
Detrended 1st Difference



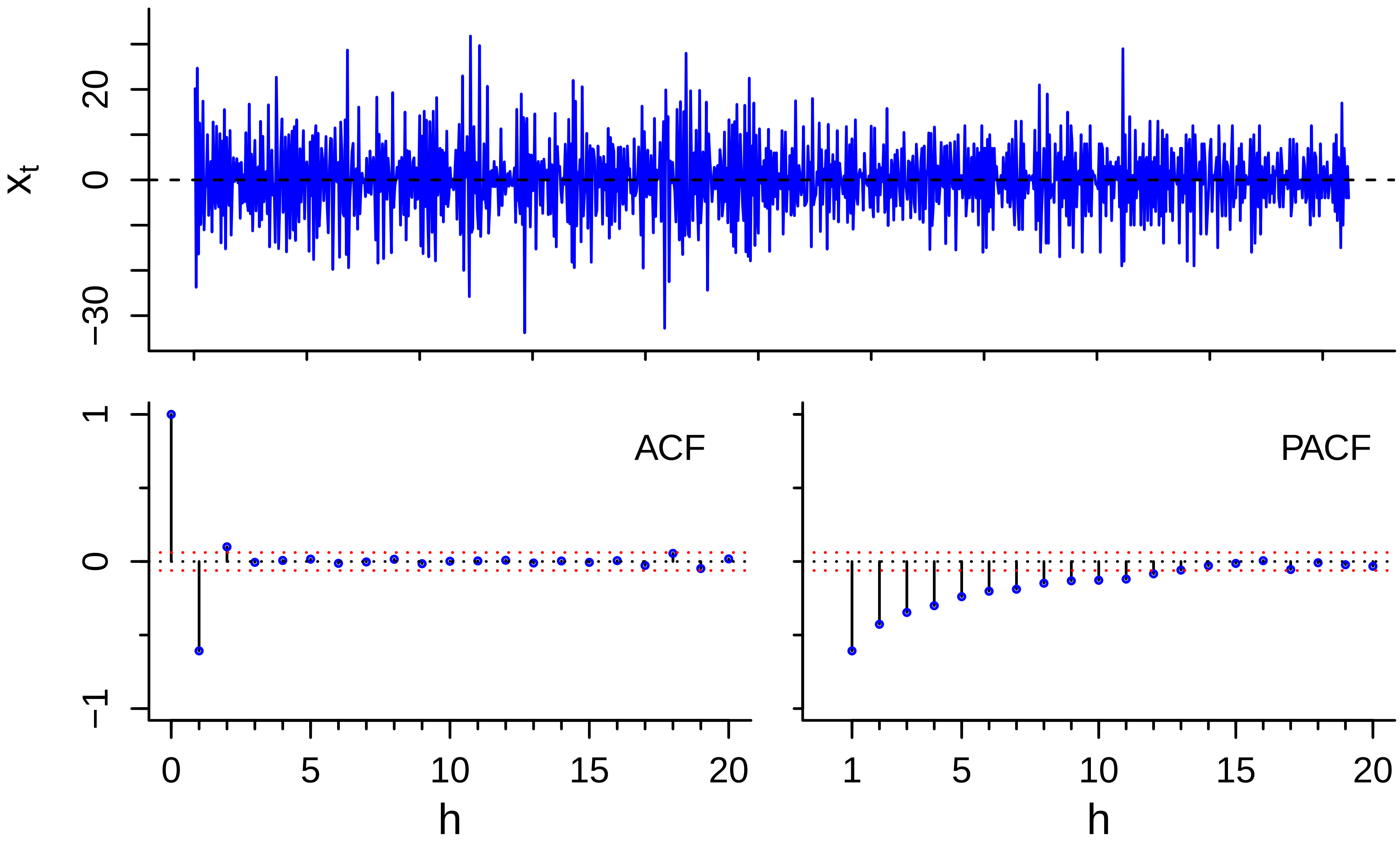
2nd Difference of Atomic Clock Time Series



2nd Difference of Random Walk Time Series: I



3rd Difference of Atomic Clock Time Series



Atomic Clock Deviates: III

- 1%, 5% and 10% percentage points are -3.43, -2.86 and -2.57

	test	$\hat{\phi}_1'$	$\widehat{\mathrm{SE}}(\hat{\phi}_1')$	t	null hypothesis
1st difference	DF	-0.10067	0.01378	-7.307	reject
1st difference	ADF	-0.02465	0.01296	-1.902	fail to reject
detrended 1st difference	DF	-0.39494	0.02481	-15.918	reject
detrended 1st difference	ADF	-0.16494	0.03173	-5.198	reject

- presuming that ADF with $p = 8$ is more appropriate than DF, test suggests $\nabla^2 X_t$ is to be preferred over ∇X_t
- found ARMA(1,1) model to adequately describe $\nabla^2 X_t$, implying ARIMA(1,2,1) model for original data:

$$\phi(B)(1 - B)^2 X_t = \theta(B) Z_t$$

Atomic Clock Deviates: IV

- note: actual model needs to include term for mean since apparent linear drift in ∇X_t would imply nonzero mean in $\nabla^2 X_t$
- unit root tests on linearly detrended series ∇X_t suggest competing model

$$X_t = a + bt + ct^2 + W_t \text{ so that } \nabla X_t = \alpha + \beta t + \nabla W_t$$

with $\alpha \stackrel{\text{def}}{=} b - c$ and $\beta \stackrel{\text{def}}{=} 2c$, where ∇W_t is an ARMA process (or possibly another stationary process)

References

- D. A. Dickey and W. A. Fuller (1979), 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root,' *Journal of the American Statistical Association*, **74**, pp. 427–431
- E. Zivot and J. Wang (2006), *Modeling Financial Time Series with S-PLUS* (Second Edition), New York: Springer