Classical Decomposition Model Revisited: I

- recall classical decomposition model for time series Y_t, namely,

$$Y_t = m_t + s_t + W_t, \qquad (*)$$

  where m_t is trend; s_t is periodic with known period s (i.e., $s_{t-s} = s_t$ for all $t \in \mathbb{Z}$) satisfying $\sum_{j=1}^{s} s_j = 0$; and W_t is a stationary process with zero mean

- m_t & s_t often taken to be deterministic
- as we have seen, SARIMA processes can model stochastic s_t, but can also handle deterministic m_t & s_t through differencing
- differencing to eliminate deterministic m_t and/or s_t can lead to undesirable overdifferencing of W_t
- alternative approach: regard $(*)$ to be a regression model with stationary errors

Classical Decomposition Model Revisited: II

- will now consider a linear regression approach in which m_t and/or s_t depend linearly on a small number of parameters
- note: nonparametric regression another option (expands upon ideas discussed earlier of using filtering operations to extract trends – see overhead III–25 and discussion following it)

Regression with Stationary Errors: I

- in context of time series analysis, standard linear regression model would take form

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{Z},$$

  where

  - $\boldsymbol{Y} = [Y_1, \ldots, Y_n]'$ is vector containing series;
  - X is an $n \times k$ design matrix whose tth row $\mathbf{x}_t'$ has values of explanatory variables for Y_t;
  - $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k]'$ is vector of regression coefficients; and
  - $\boldsymbol{Z} = [Z_1, \ldots, Z_n]'$ is vector of $\mathrm{WN}(0, \sigma^2)$ RVs

- example: $Y_t = \beta_1 + \beta_2 t + Z_t$, for which tth row of $n \times 2$ design matrix X would be $\mathbf{x}_t' = [1, t]$
- second example: $Y_t = \beta_1 + \beta_2 \cos(2\pi f t) + \beta_3 \sin(2\pi f t) + Z_t$

Regression with Stationary Errors: II

- ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ is vector $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ minimizing sum of squared errors:

$$S(\boldsymbol{\beta}) = \sum_{t=1}^{n} (Y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2 = (\boldsymbol{Y} - X\boldsymbol{\beta})'(\boldsymbol{Y} - X\boldsymbol{\beta})$$

- $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is solution to so-called normal equations:

$$X'X\boldsymbol{\beta} = X'\boldsymbol{Y} \ \text{ and hence } \ \hat{\boldsymbol{\beta}}_{\text{OLS}} = (X'X)^{-1}X'\boldsymbol{Y}$$

  if $X'X$ has full rank
- $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is best linear unbiased estimator of $\boldsymbol{\beta}$ (best in that, if $\hat{\boldsymbol{\beta}}$ is any other unbiased estimator, then $\text{var}\{\boldsymbol{c}'\hat{\boldsymbol{\beta}}_{\text{OLS}}\} \leq \text{var}\{\boldsymbol{c}'\hat{\boldsymbol{\beta}}\}$ for any vector $\boldsymbol{c}$ of constants)
- covariance matrix for $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ given by $\sigma^2(X'X)^{-1}$

Regression with Stationary Errors: III

- for time series, uncorrelated errors $\boldsymbol{Z}$ are usually unrealistic
- often a more realistic model is

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{W},$$

  where $\boldsymbol{W}$ contains RVs from a stationary process with zero mean, an example being a causal ARMA(p, q) process:

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

- under this alternative model, $\hat{\boldsymbol{\beta}}_{\texttt{OLS}}$ is still an unbiased estimator of $\boldsymbol{\beta}$, but best linear unbiased estimator is generalized least squares (GLS) estimator
- in the full rank case, this estimator takes the form

$$\hat{\boldsymbol{\beta}}_{\texttt{GLS}} = (X'\Gamma_n^{-1}X)^{-1}X'\Gamma_n^{-1}\boldsymbol{Y},$$

  where Γ_n is covariance matrix for $\boldsymbol{W}$

Regression with Stationary Errors: IV

- $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ is minimizer of *weighted* sum of squares:

$$S(\boldsymbol{\beta}) = (\boldsymbol{Y} - X\boldsymbol{\beta})'\Gamma_n^{-1}(\boldsymbol{Y} - X\boldsymbol{\beta})$$

- to motivate this estimator, recall innovations representation for stationary process says that $\boldsymbol{W} = C_n\boldsymbol{U}$, where $\boldsymbol{U} = \boldsymbol{W} - \widehat{\boldsymbol{W}}$ ($\boldsymbol{U}$ are the innovations, and $\widehat{\boldsymbol{W}}$ are 1-step-ahead predictions)
- recall that, if $\boldsymbol{V}$ has covariance matrix Σ_n, then covariance matrix for $A\boldsymbol{V}$ is $A\Sigma_n A'$
- since $\boldsymbol{U}$ has diagonal covariance matrix D_n, get $\Gamma_n = C_n D_n C_n'$
- let $D_n^{-1/2}$ be diagonal matrix such that $D_n^{-1/2}D_n^{-1/2} = D_n^{-1}$
- covariance matrix for $D_n^{-1/2}C_n^{-1}\boldsymbol{W}$ is identity matrix I_n since

$$D_n^{-1/2}C_n^{-1}\Gamma_n(C_n^{-1})'D_n^{-1/2} = D_n^{-1/2}C_n^{-1}C_n D_n C_n'(C_n^{-1})'D_n^{-1/2} = I_n$$

Regression with Stationary Errors: V

- returning now to model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{W}$, multiplication of both sides by $D_n^{-1/2}C_n^{-1}$ yields

$$D_n^{-1/2}C_n^{-1}\boldsymbol{Y} = D_n^{-1/2}C_n^{-1}X\boldsymbol{\beta} + D_n^{-1/2}C_n^{-1}\boldsymbol{W},$$

which can be reexpressed as

$$\widetilde{\boldsymbol{Y}} = \widetilde{X}\boldsymbol{\beta} + \boldsymbol{Z}, \ \ \{Z_t\} \sim \mathrm{WN}(0,1)$$

- for this regression model, best linear unbiased estimator is

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_{\texttt{OLS}} &= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{\boldsymbol{Y}} \\
&= (X'(C_n^{-1})'D_n^{-1/2}D_n^{-1/2}C_n^{-1}X)^{-1}X'(C_n^{-1})'D_n^{-1/2}D_n^{-1/2}C_n^{-1}\boldsymbol{Y} \\
&= (X'(C_n^{-1})'D_n^{-1}C_n^{-1}X)^{-1}X'(C_n^{-1})'D_n^{-1}C_n^{-1}\boldsymbol{Y} \\
&= (X'\Gamma_n^{-1}X)^{-1}X'\Gamma_n^{-1}\boldsymbol{Y} = \hat{\boldsymbol{\beta}}_{\texttt{GLS}}
\end{aligned}$$

since $\Gamma_n = C_nD_nC_n'$ says $\Gamma_n^{-1} = (C_n')^{-1}D_nC_n^{-1} = (C_n^{-1})'D_nC_n^{-1}$

Regression with Stationary Errors: VI

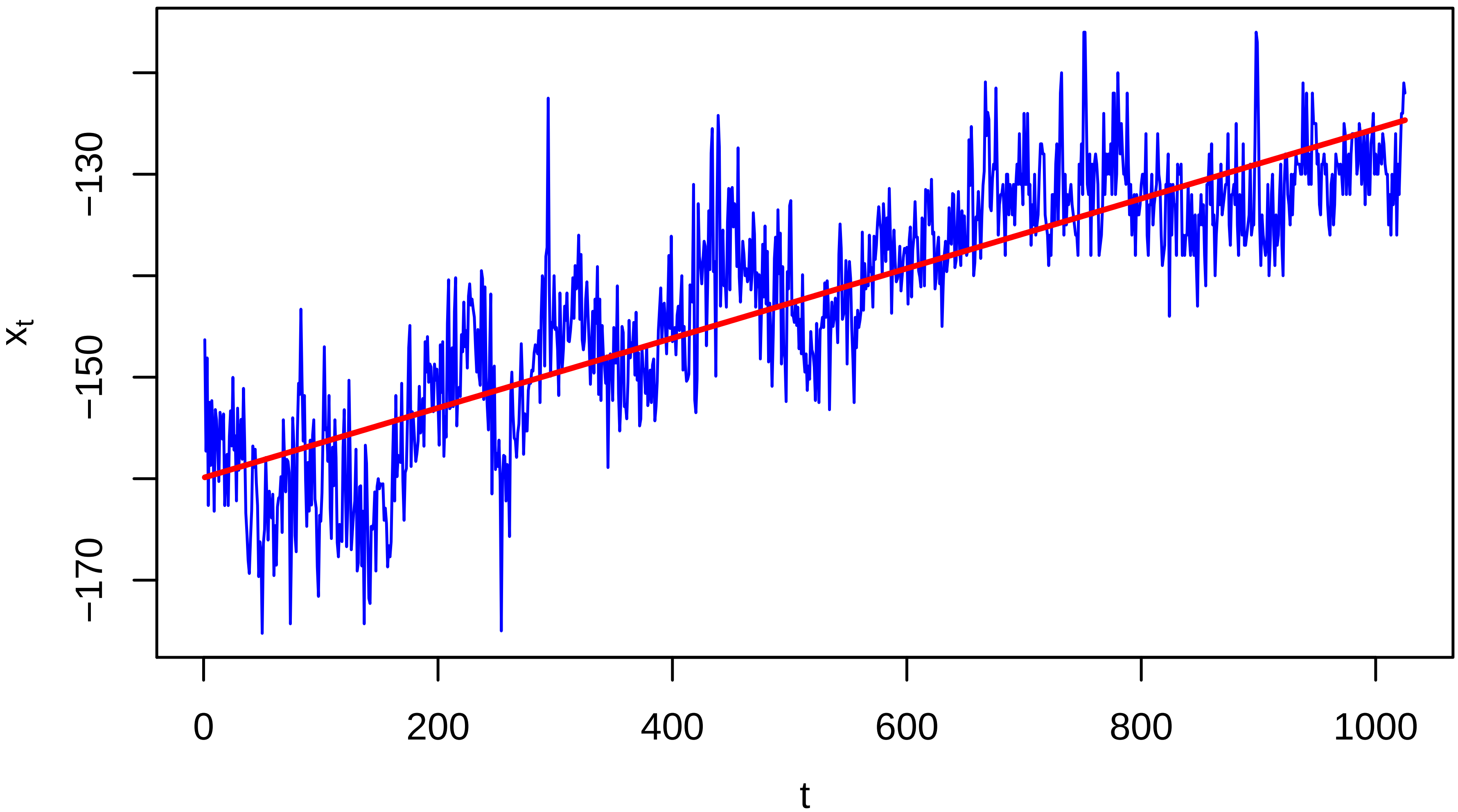
- in principle, can use ML under a Gaussian assumption to estimate all parameters in model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{W}$ (i.e., both $\boldsymbol{\beta}$ and parameters associated with, e.g., ARMA model)
- in practice, following simpler (but sub-optimal) iterative scheme for parameter estimation often works well

1. compute $\hat{\boldsymbol{\beta}}_{\texttt{OLS}}$ and form residuals $Y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{\texttt{OLS}}$
2. fit ARMA(p,q) or other stationary model to residuals
3. using fitted model, compute $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ and form residuals $Y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$
4. fit same model to residuals again
5. repeat steps 3 and 4 until parameter estimates have stabilized

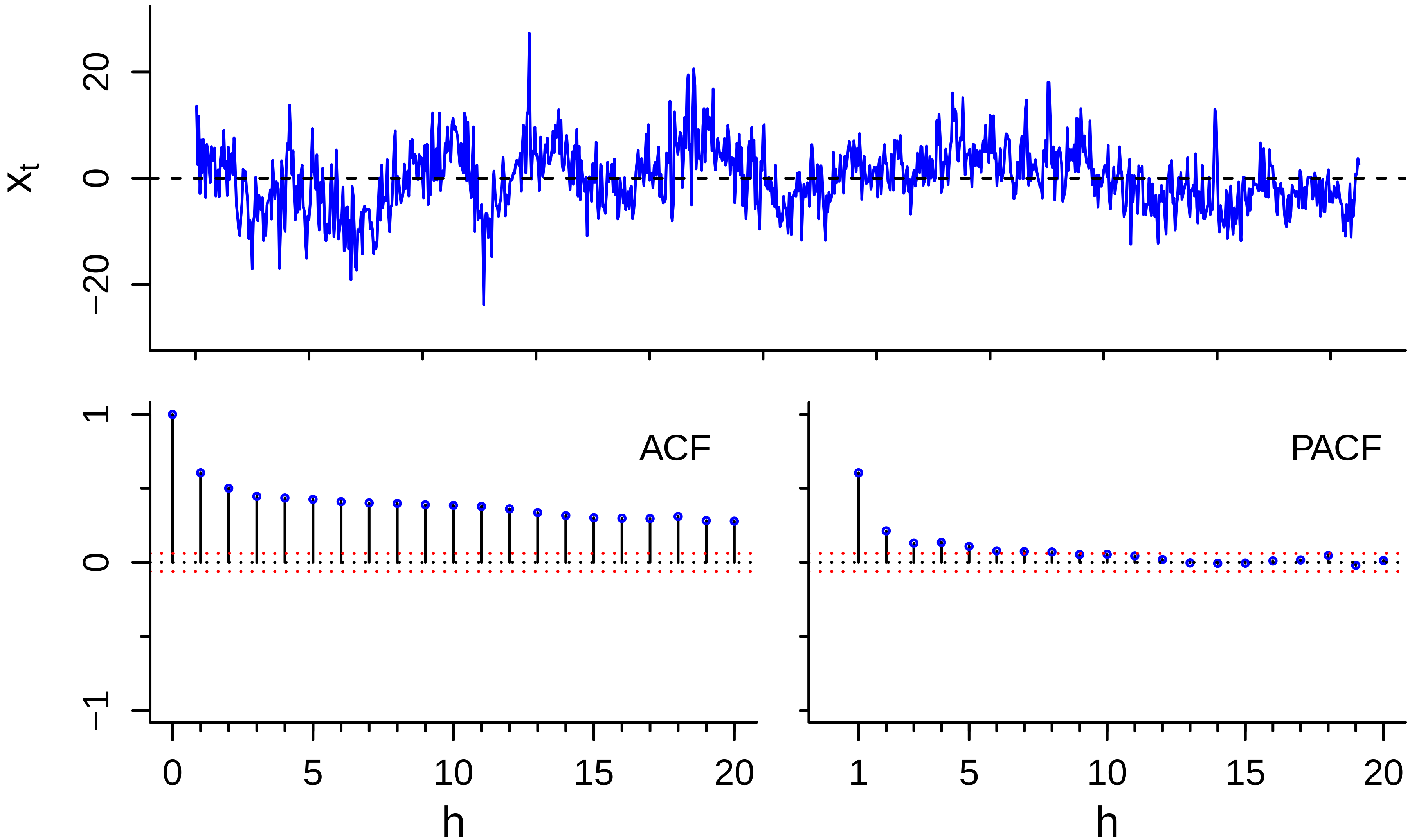
1st Difference of Atomic Clock Data Revisited: I

- let's reconsider modeling 1st difference of atomic clock data X_t, i.e., ∇X_t
- 2nd difference $\nabla^2 X_t$ well-modeled by ARMA(1,1) process (overheads XIII–129 to 135), with support for need for 2nd differencing coming from augmented Dickey–Fuller (ADF) unit root test (overhead XIV–44)
- modeling ∇X_t as a stationary fractionally differenced (FD) process proved to be questionable (overheads XVI–72 to 77)
- apparent linear increase in ∇X_t suggests exploring model of linear trend + stationary noise more seriously (looked briefly at this approach in context of ADF unit root test, where we found that linear detrending obviated need for 2nd differencing – see overheads XIV–44 and 45)

1st Difference of Atomic Clock Time Series



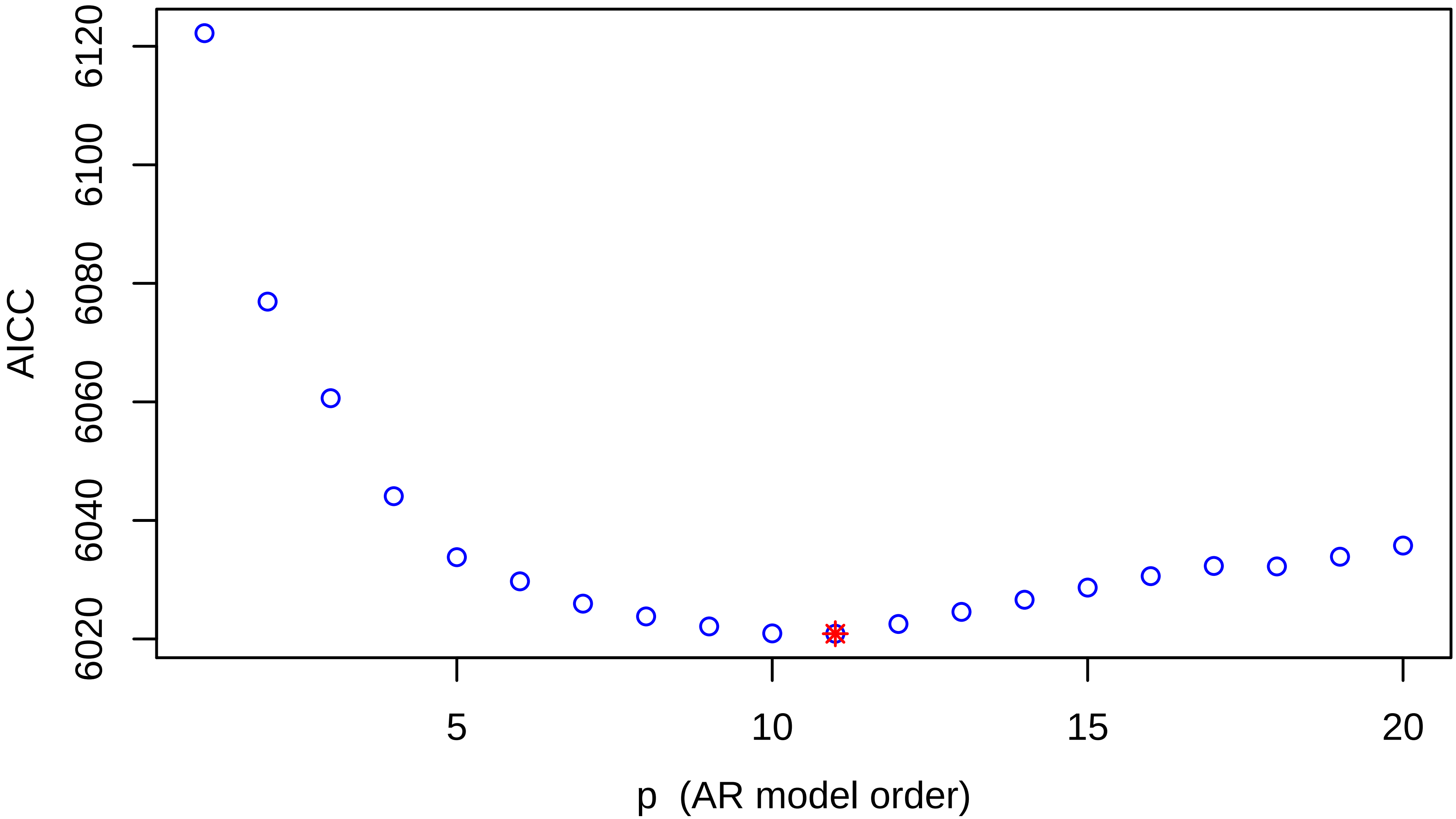
Residuals from OLS Fit of Linear Regression



1st Difference of Atomic Clock Data Revisited: II

- sample ACF and PACF suggest three models for residuals:
  - AR(p), with p somewhere around 7 or 8
  - ARMA(p,q), with $p + q$ small
  - ARFIMA(p,δ,q), with $p + q$ small ($= 0$ gives FD(δ))
- let's consider various possibilities, using
  - maximum likelihood to fit each model
  - AICC to evaluate different models
- start by considering AR(p) models of orders $p = 1, \ldots, 20$

ML-Based AICC for Residuals Modeled by AR(p)



1st Difference of Atomic Clock Data Revisited: III

- AR(11) best amongst AR(p) with AICC of 6020.89
- here are AICCs for ARMA(p,q) models with $p + q \leq 4$

p	q	AICC
0	1	6301.19
0	2	6197.77
0	3	6159.83
0	4	6135.89
1	1	6033.78
1	2	6010.36
1	3	6008.19
2	1	**6007.07**
2	2	6009.01
3	1	6008.98

1st Difference of Atomic Clock Data Revisited: IV

- let's compare AICC for ARMA(2,1) with ones for ARMA(p,q) models with $p + q = 5$

p	q	AICC
2	1	**6007.07**
0	5	6114.91
1	4	6010.18
2	3	6010.19
3	2	6010.71
4	1	6010.01

- let's now see if we can find an ARFIMA(p,δ,q) model that has a better AICC than one for ARMA(2,1) model

1st Difference of Atomic Clock Data Revisited: V

p	δ	q	AICC
2	0	1	**6007.07**
0	0.405	0	6015.17
0	0.391	1	6015.10
0	0.426	2	6015.45
0	0.430	3	6015.57
1	0.395	0	6015.14
1	0.398	1	6008.70
1	0.097	2	6007.30
2	0.417	0	6015.85
2	0.000	1	6007.61
3	0.432	0	6015.08

- conclusion: will go with ARMA(2,1) model

1st Difference of Atomic Clock Data Revisited: VI

- using `R` function **`arima`** to fit model

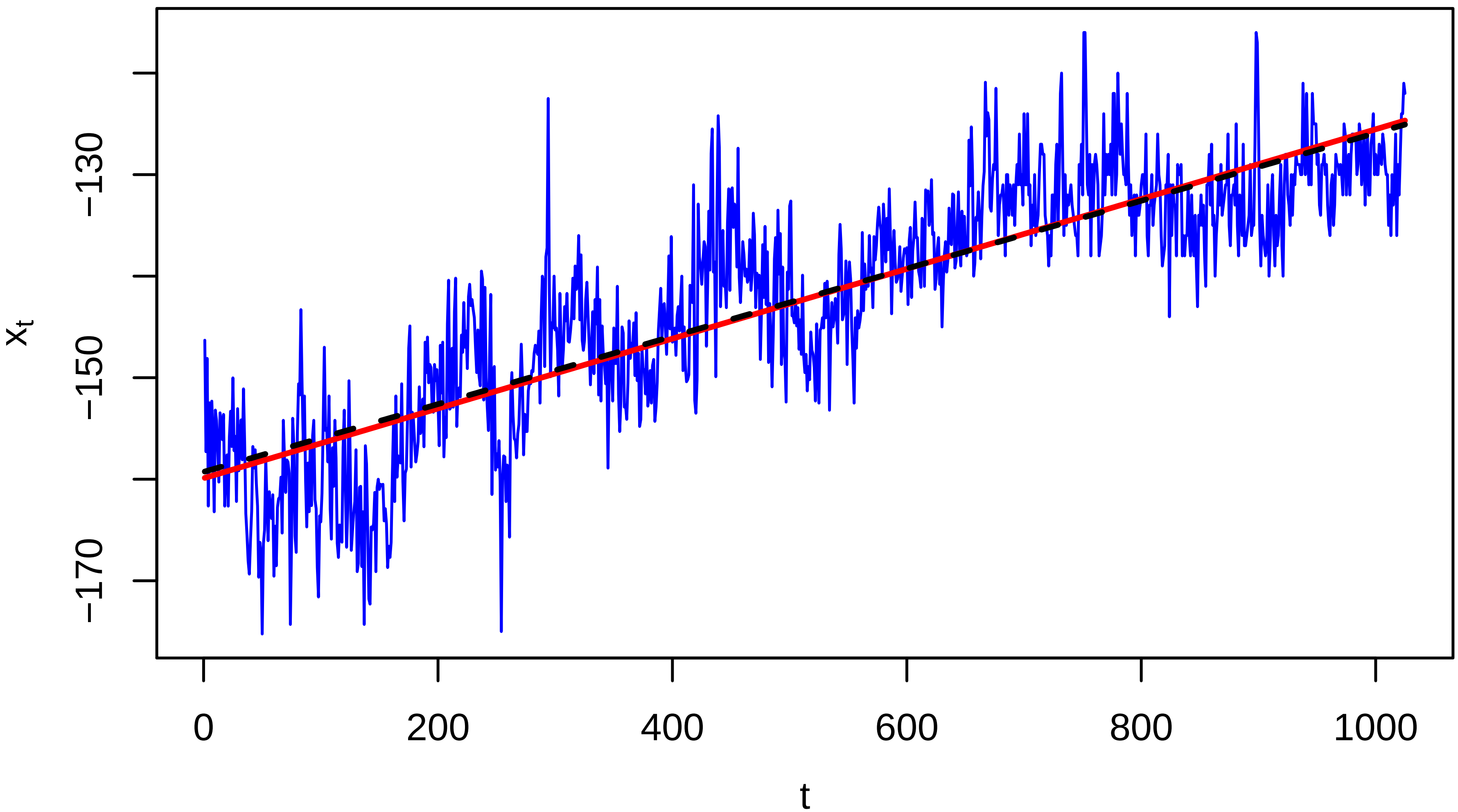
$$\nabla X_t = a + bt + W_t, \ \ \{W_t\} \sim \text{ARMA}(2{,}1),$$

ML estimates of model parameters a, b, ϕ_1, ϕ_2 and θ_1 (along with standard errors (SEs)) are

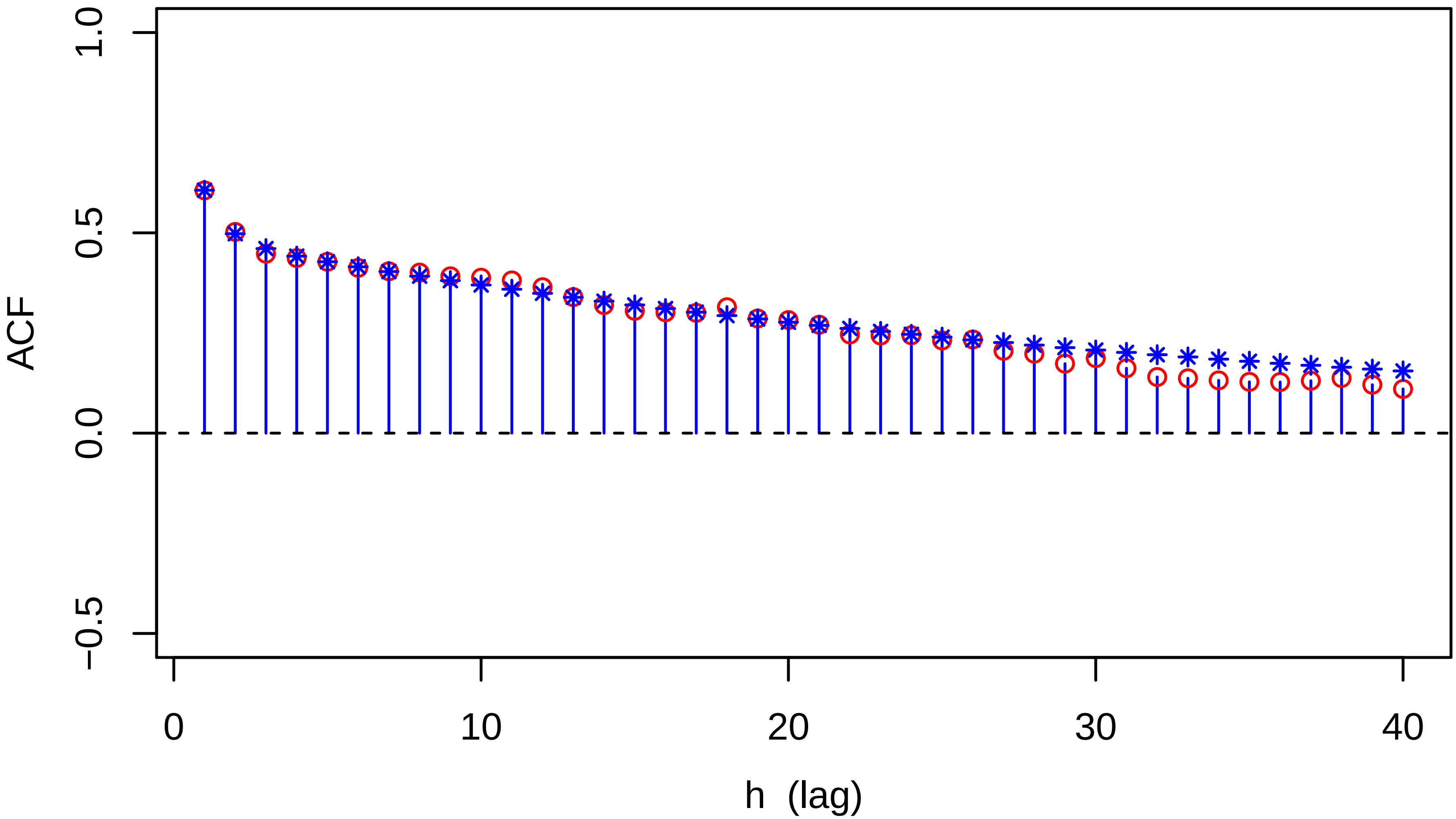
	$\hat{a}$	$\hat{b}$	$\hat{\phi}_1$	$\hat{\phi}_1$	$\hat{\theta}_1$
value	-159.2759	0.0334	1.2218	-0.2432	-0.8315
SE	2.0863	0.0035	0.0453	0.0412	0.0299

- estimate of σ^2 is $\hat{\sigma}^2 \doteq 20.3653$; next plots show
  - ∇X_t & lines fitted by OLS (red) & GLS (black dashed)
  - comparison of ARMA(2,1) ACF & PACF with sample ACF & PACF for $\nabla X_t - \hat{a} - \hat{b}t$
  - residuals $\hat{R}_t$ returned by **`arima`** and associated diagnostics

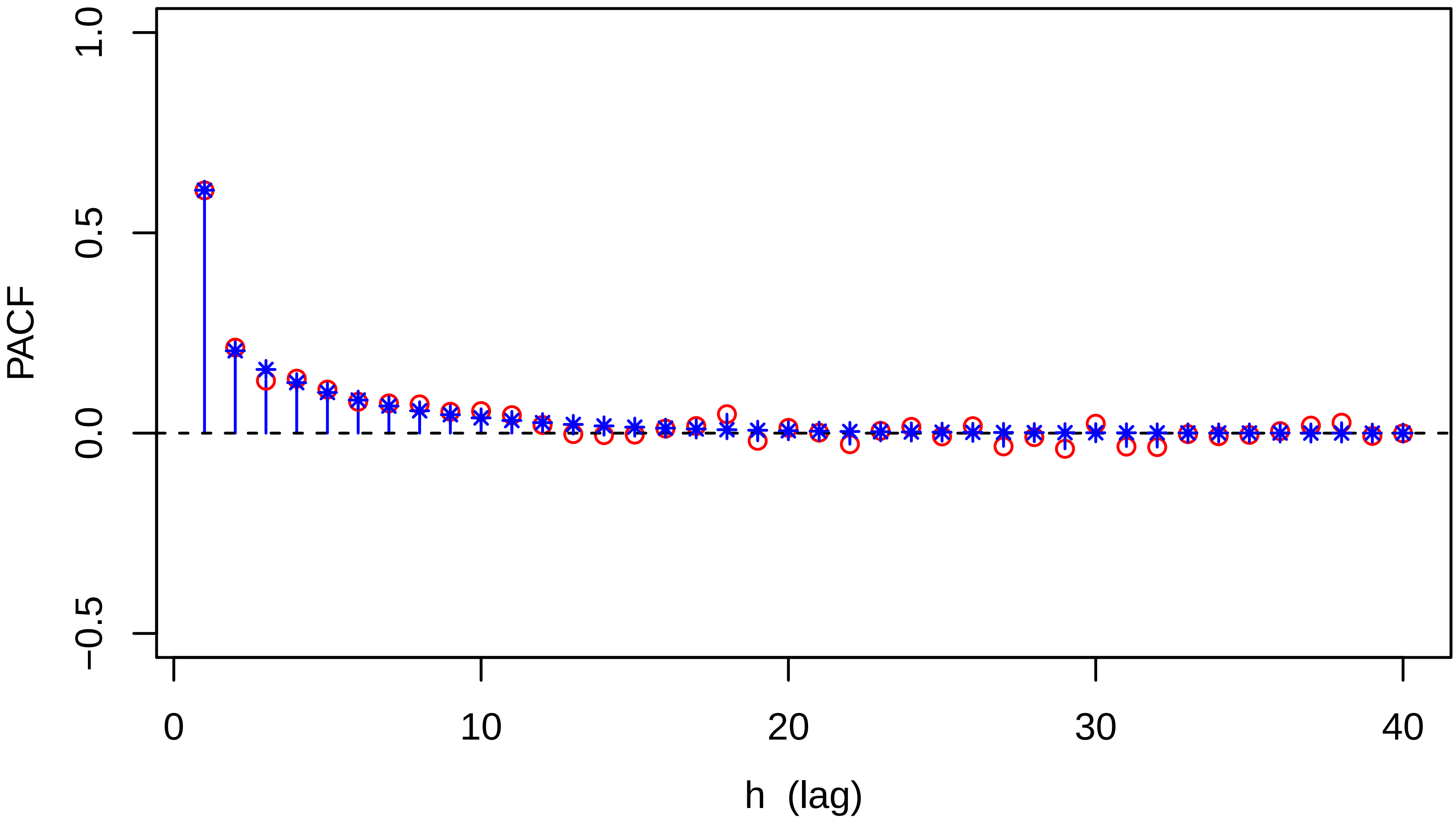
1st Difference of Atomic Clock Time Series



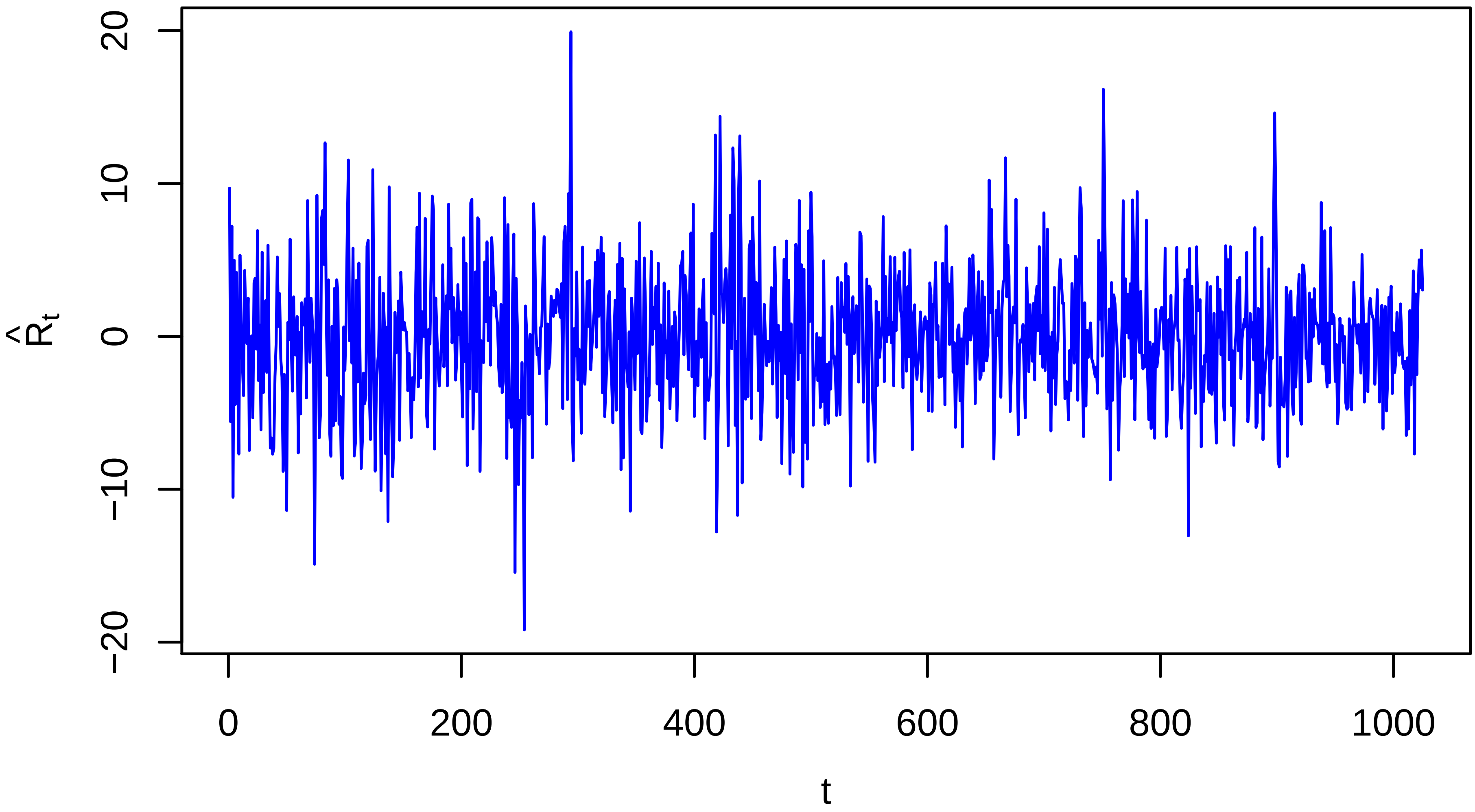
Sample ACF for $\nabla X_t - \hat{a} - \hat{b}t$ & ARMA(2,1) ACF



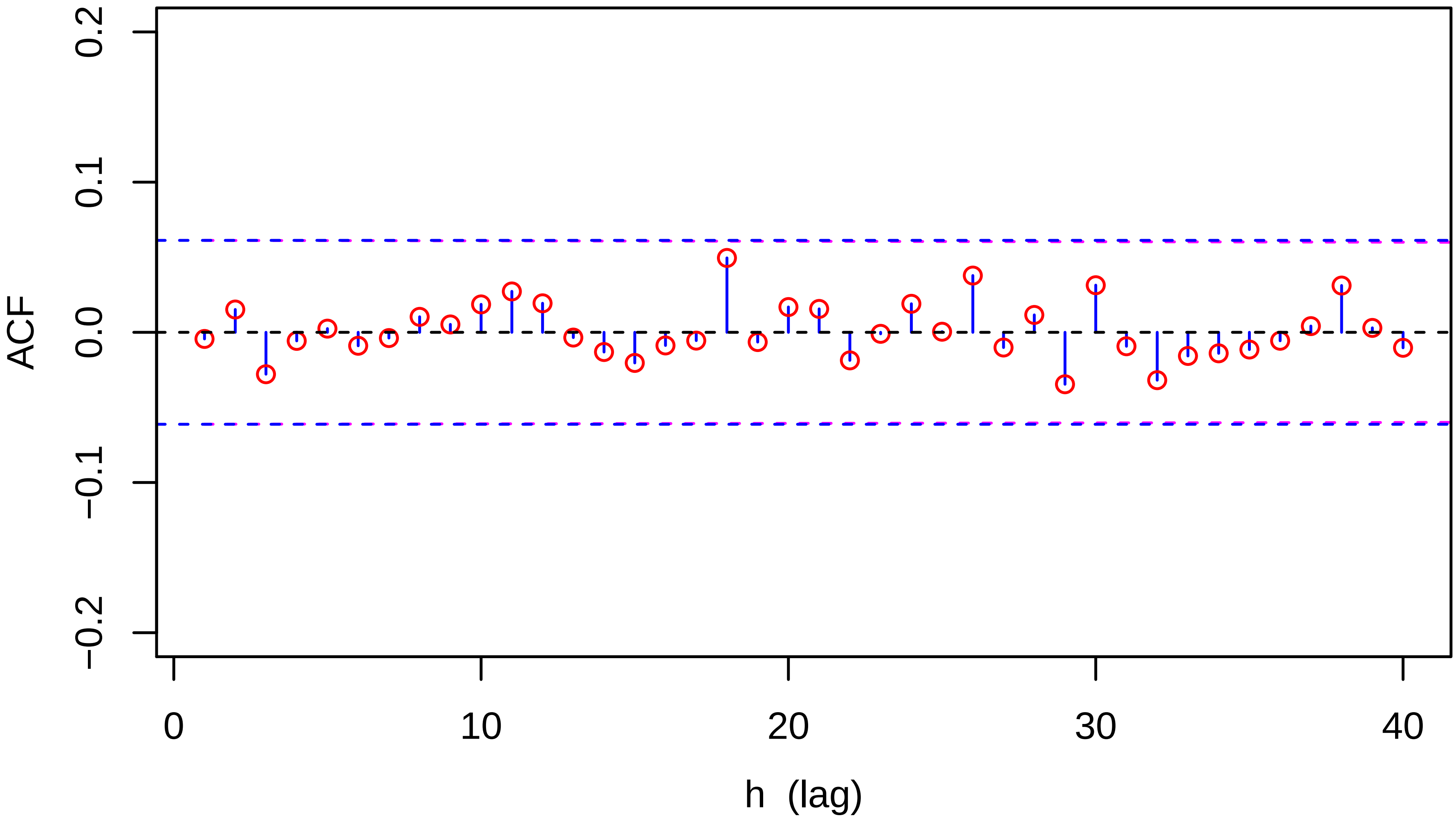
Sample PACF for $\nabla X_t - \hat{a} - \hat{b}t$ & ARMA(2,1) PACF



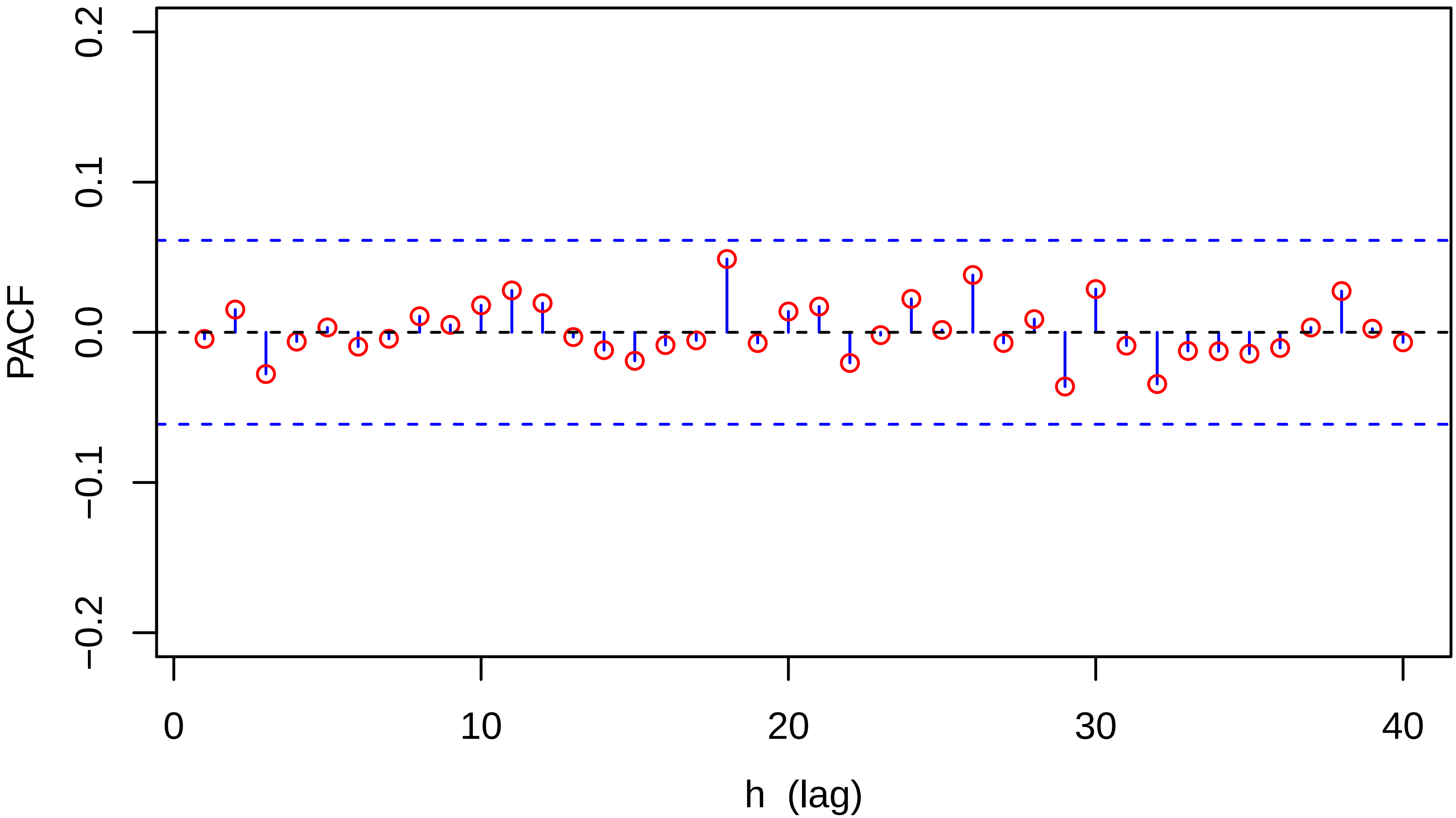
ARMA(2,1) Residuals $\hat{R}_t$ for Atomic Clock



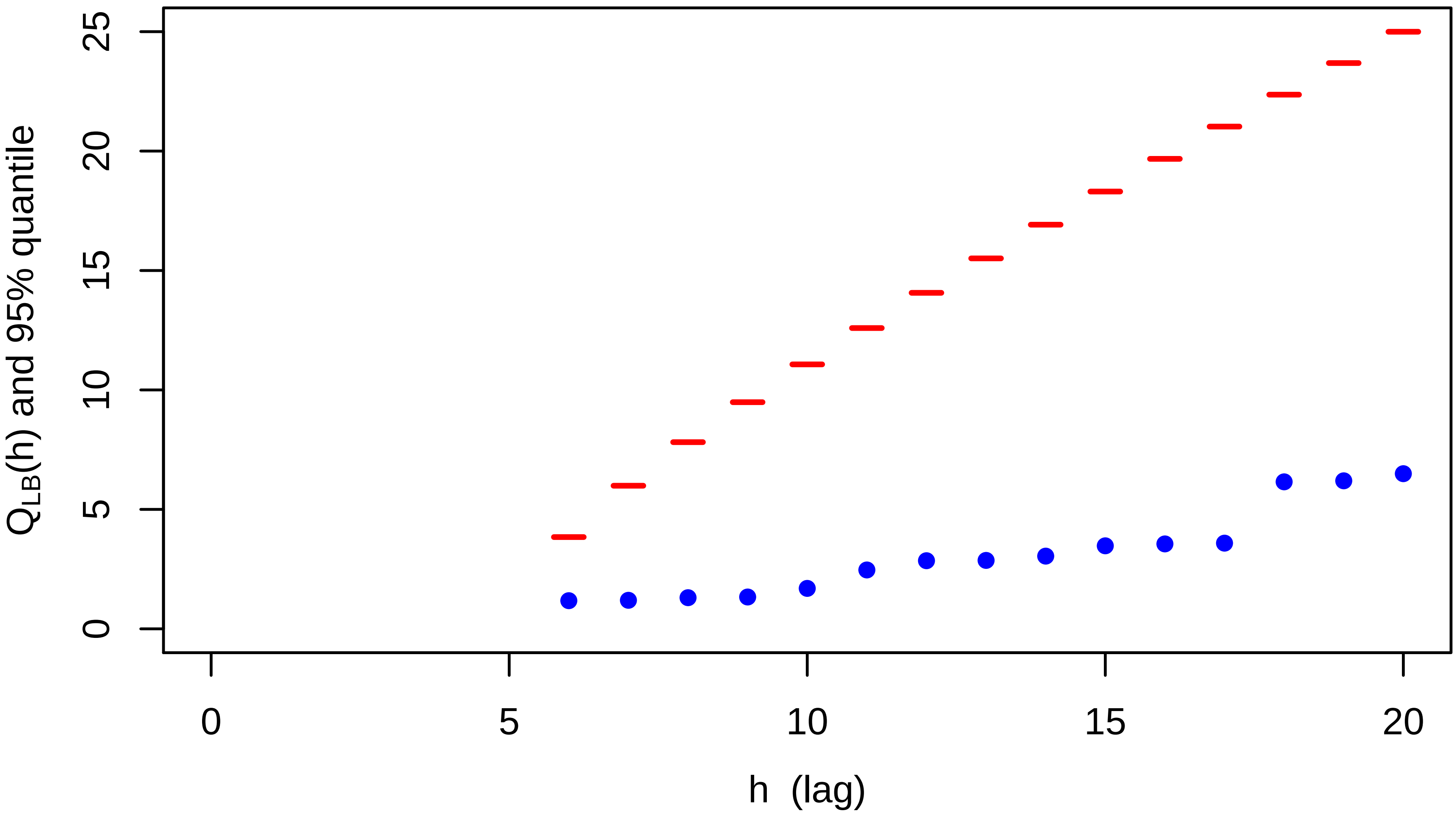
Sample ACF for ARMA(2,1) Residuals $\hat{R}_t$



Sample PACF for ARMA(2,1) Residuals $\hat{R}_t$



Portmanteau Tests of ARMA(2,1) Residuals $\hat{R}_t$



Diagnostics Tests of ARMA(2,1) Residuals $\hat{R}_t$

test	expected value	test statistic	p-value
turning point	682	708	**0.054**
difference-sign	512	512	1.00
rank	262400	263233	0.879
runs	513.2	525	0.460

AR method	AICC order	AIC order
Yule–Walker	0	0
Burg	0	0
OLS	0	**4**
MLE	0	0

Regression with Stationary Errors: VII

- cautionary note: $\boldsymbol{Y}'\boldsymbol{Y} \neq \widetilde{\boldsymbol{Y}}'\widetilde{\boldsymbol{Y}} = \boldsymbol{Y}'\Gamma_n^{-1}\boldsymbol{Y}$, so portion of sum of squares explained by transformed model cannot be related directly to sum of squares for untransformed data
- assuming $\boldsymbol{W}$ in model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{W}$ has covariance matrix Γ_n, covariance matrices for $\hat{\boldsymbol{\beta}}_{\mathtt{OLS}}$ and $\hat{\boldsymbol{\beta}}_{\mathtt{GLS}}$ are, respectively,
$$(X'X)^{-1}X'\Gamma_n X(X'X)^{-1} \text{ and } (X'\Gamma_n^{-1}X)^{-1}$$
- assuming $\boldsymbol{W}$ for ∇X_t clock data is ARMA(2,1) process, can assess standard errors (SEs) for estimated slopes

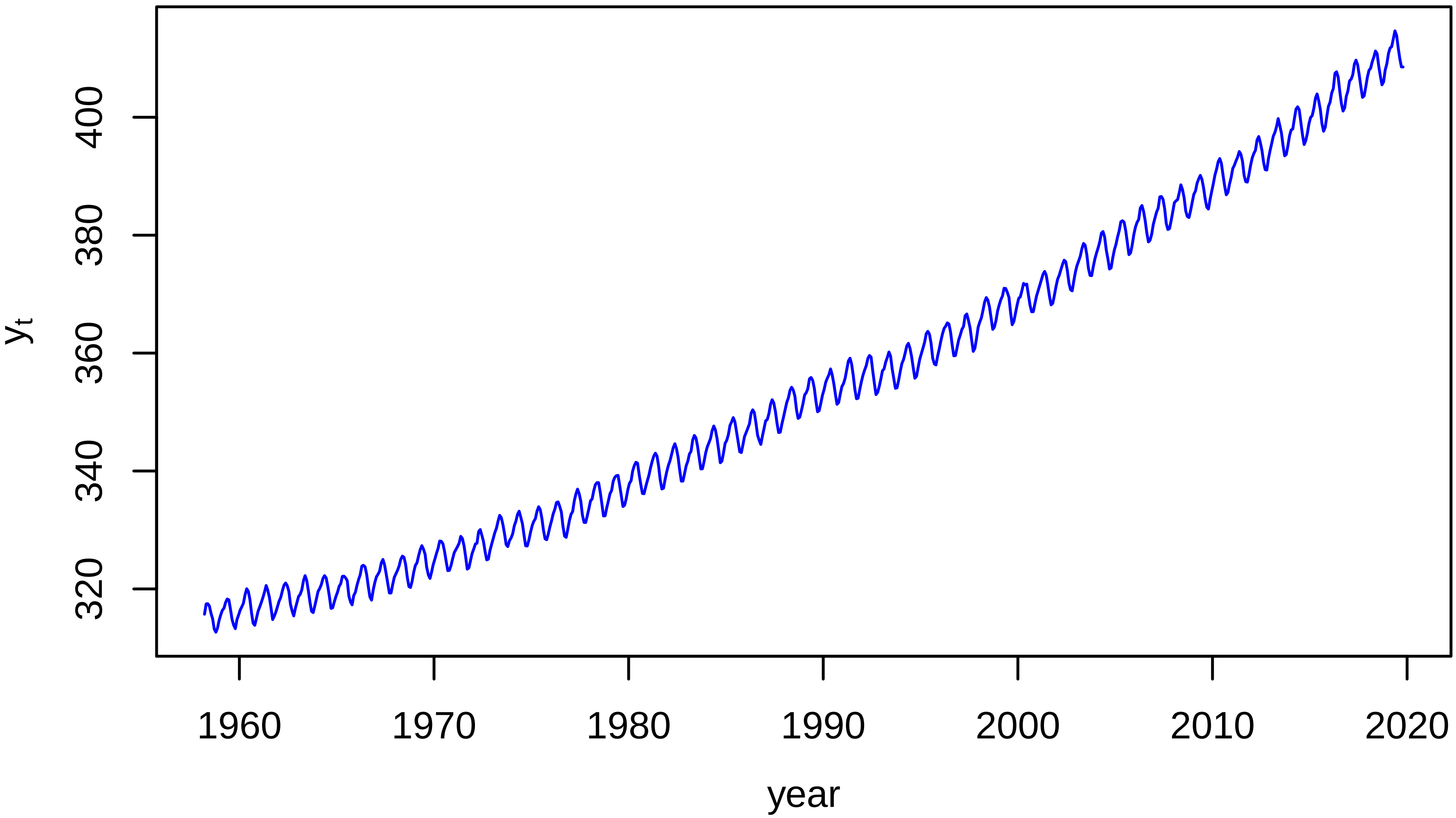
	OLS	GLS	
slope	0.03437	0.03337	($\approx$ 3% difference between estimates)
SE	0.00358	0.00346	($\approx$ 3% increase for OLS over GLS)

- if assume $\boldsymbol{W}$ is WN, SE for OLS slope estimate would be taken as 0.00064 (5× smaller than when correlation is accounted for)

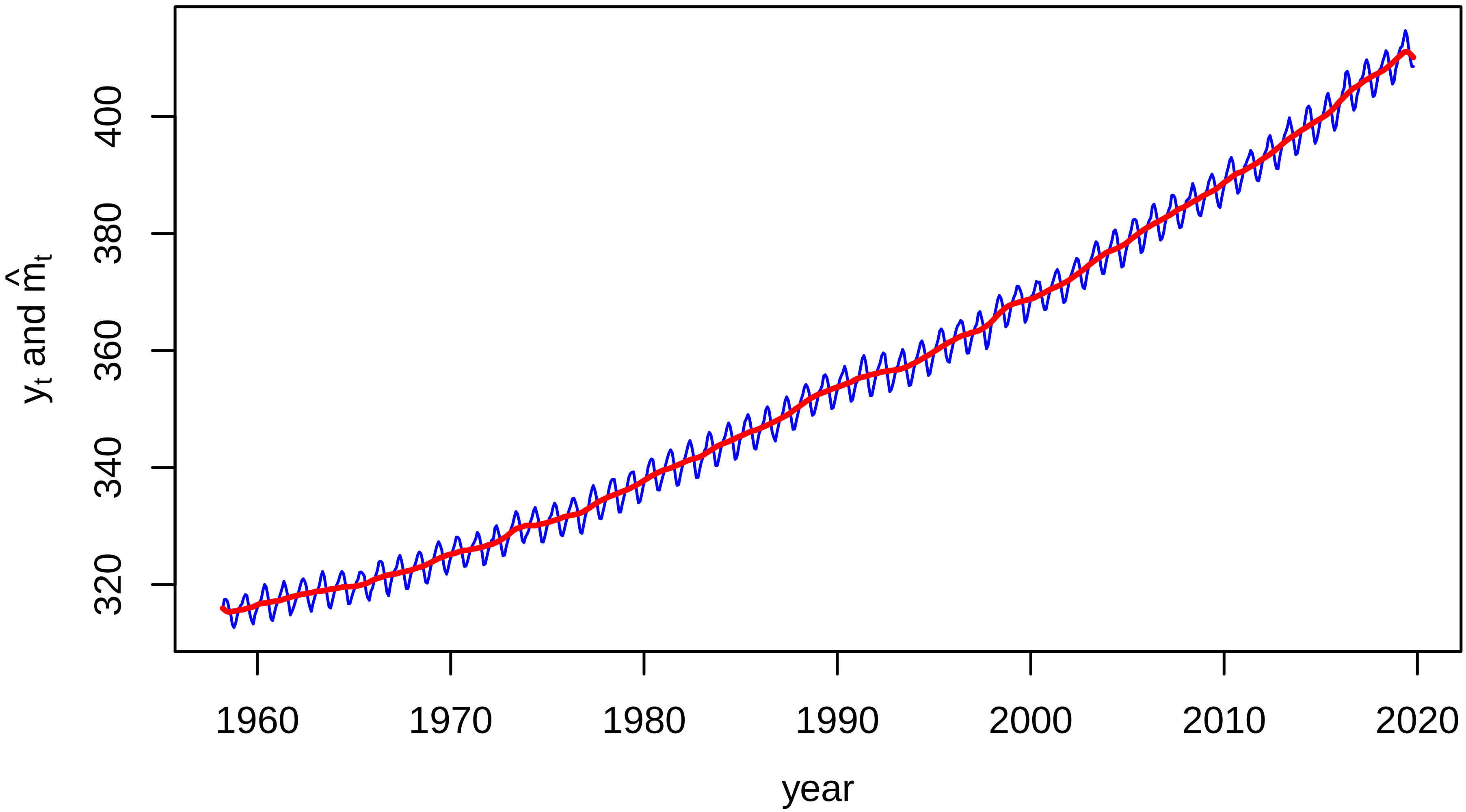
Harmonic Regression and CO_2 Series: I

- atomic clock ∇X_t illustrates handling deterministic trend m_t in model
$$Y_t = m_t + W_t$$
via a parametric regression approach ($m_t = a + bt$)
- as a second example, reconsider CO_2 series from Mauna Loa, Hawaii (subject of Problem 8)
- appropriate model here is full classical decomposition model:
$$Y_t = m_t + s_t + W_t$$
where m_t is trend; s_t is seasonal component with period $s = 12$ (i.e., $s_{t-12} = s_t$ for all $t \in \mathbb{Z}$) satisfying $\sum_{j=1}^{12} s_j = 0$; and W_t is a stationary process with zero mean
- for illustrative purposes, will take m_t & s_t to be deterministic

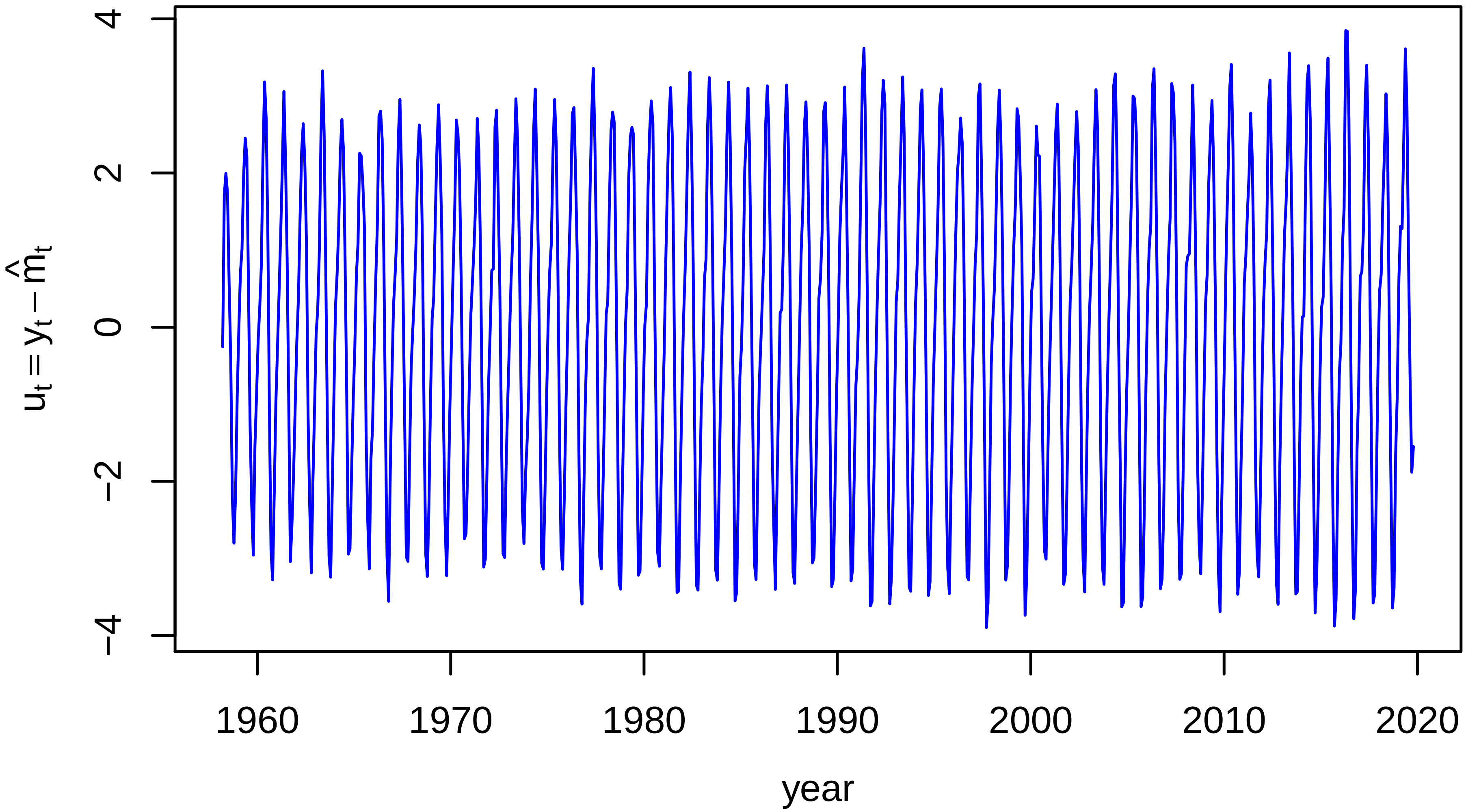
2nd Example: CO_2 Series from Mauna Loa, Hawaii

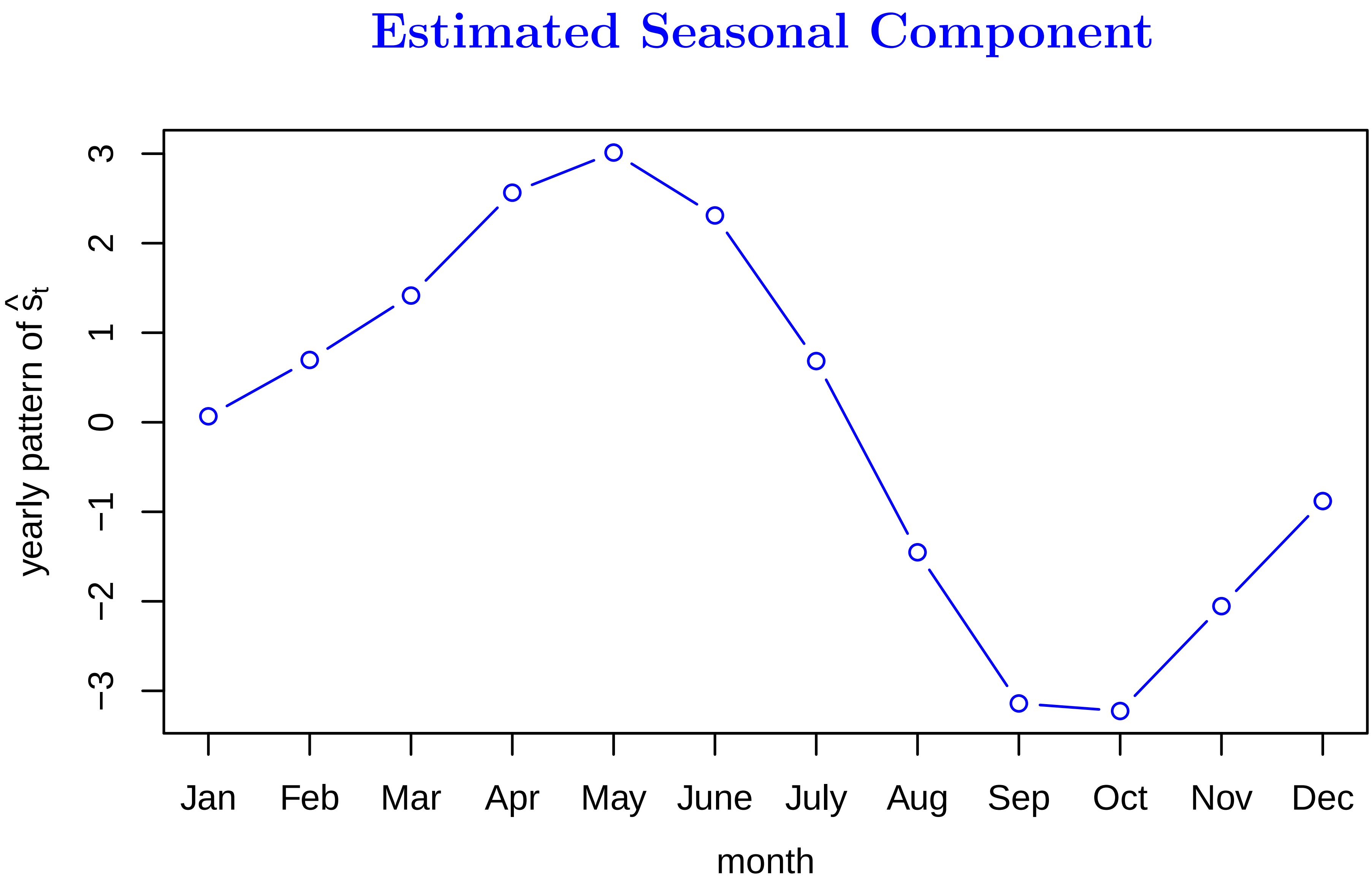


Preliminary Nonparametric Estimate of Trend



Preliminary Detrended Series




Estimated Seasonal Component
yearly pattern of $\hat{s}_t$
3
2
1
0
−1
−2
−3
Jan
Feb
Mar
Apr
May
June
July
Aug
Sep
Oct
Nov
Dec
month

Harmonic Regression and CO_2 Series: II

- factoid: can represent *any* deterministic s_t with period 12 by

$$s_t = \sum_{j=1}^{6} A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t) = \sum_{j=1}^{6} D_j \cos(2\pi f_j t + \varphi_j),$$

  where $f_j \stackrel{\text{def}}{=} j/12$ is a frequency with associated period $1/f_j$: treating t momentarily as $t \in \mathbb{R}$ rather than $t \in \mathbb{Z}$, have

$$\cos(2\pi f_j(t + \tfrac{1}{f_j})) = \cos(2\pi f_j t + 2\pi) = \cos(2\pi f_j t)$$

  with a similar result holding for $\sin(2\pi f_j t)$.

- since $\sum_{j=1}^{12} s_j = 0$, any eleven s_j's determine remaining twelfth
- six A_j's and six B_j's seem one too many, but in fact there are only five relevant B_j's: B_6 doesn't enter in play because

$$B_6 \sin(2\pi f_6 t) = B_6 \sin(2\pi \tfrac{6}{12} t) = B_6 \sin(\pi t) = 0 \text{ for all } t$$

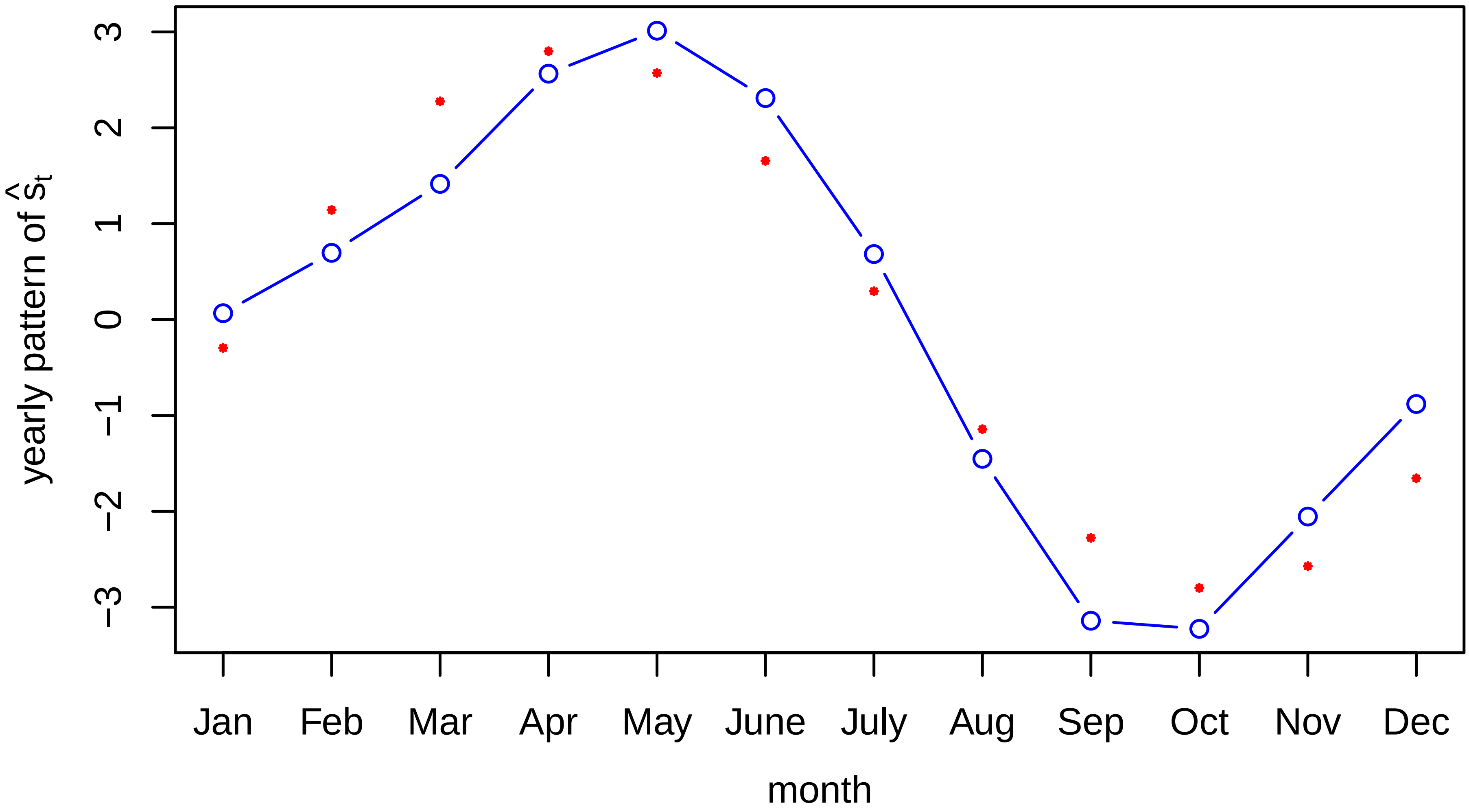
Harmonic Regression and CO_2 Series: III

- f_1 is known as the fundamental frequency, whereas $f_2, \ldots, f_6$ are called first, …, fifth harmonics
- since $f_j = j/12$, have $f_j = jf_1$ for $j = 2, \ldots, 6$, so harmonics are integer multiples of fundamental frequency
- if seasonal component s_t slowly varying from one month to next, can often get by with using just fundamental frequency and a small number of harmonics:

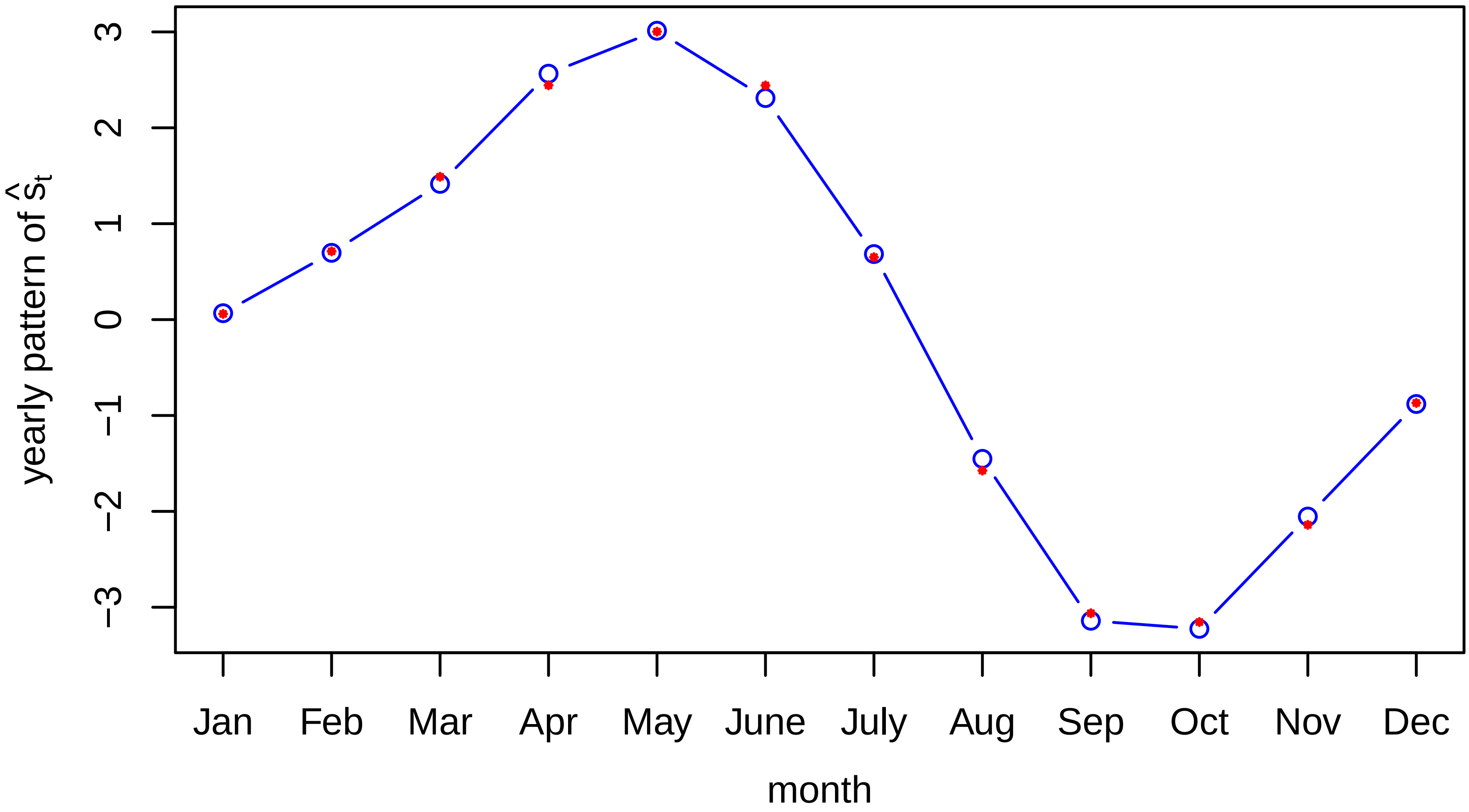
$$s_t \approx \sum_{j=1}^{J} A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t), \quad 1 \leq J < 6$$

- following plots show approximations of orders $J = 1, \ldots 5$ to seasonal component $\hat{s}_t$ estimated in Problem 8 for CO_2 series

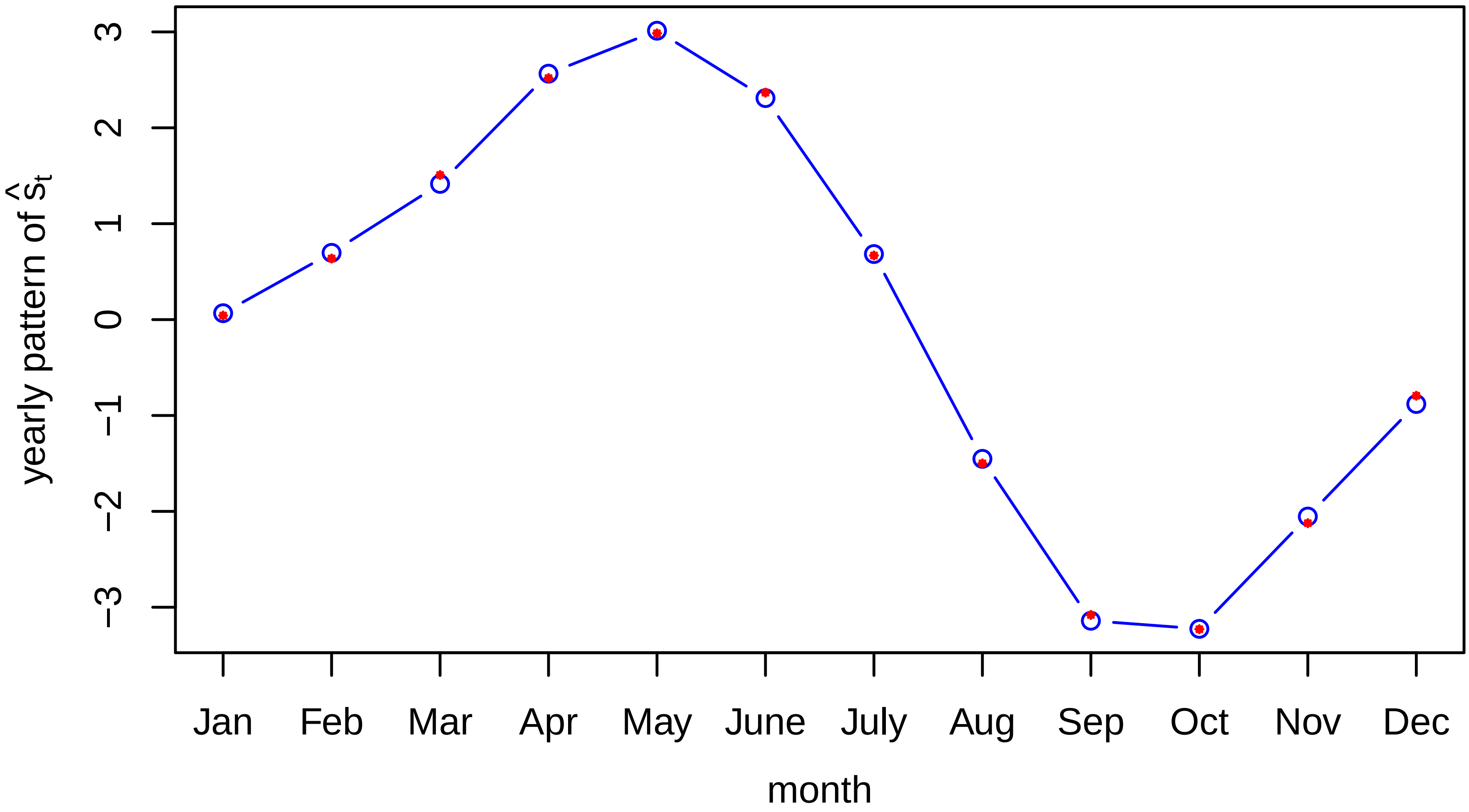
$J = 1$ Approximation to Seasonal Component



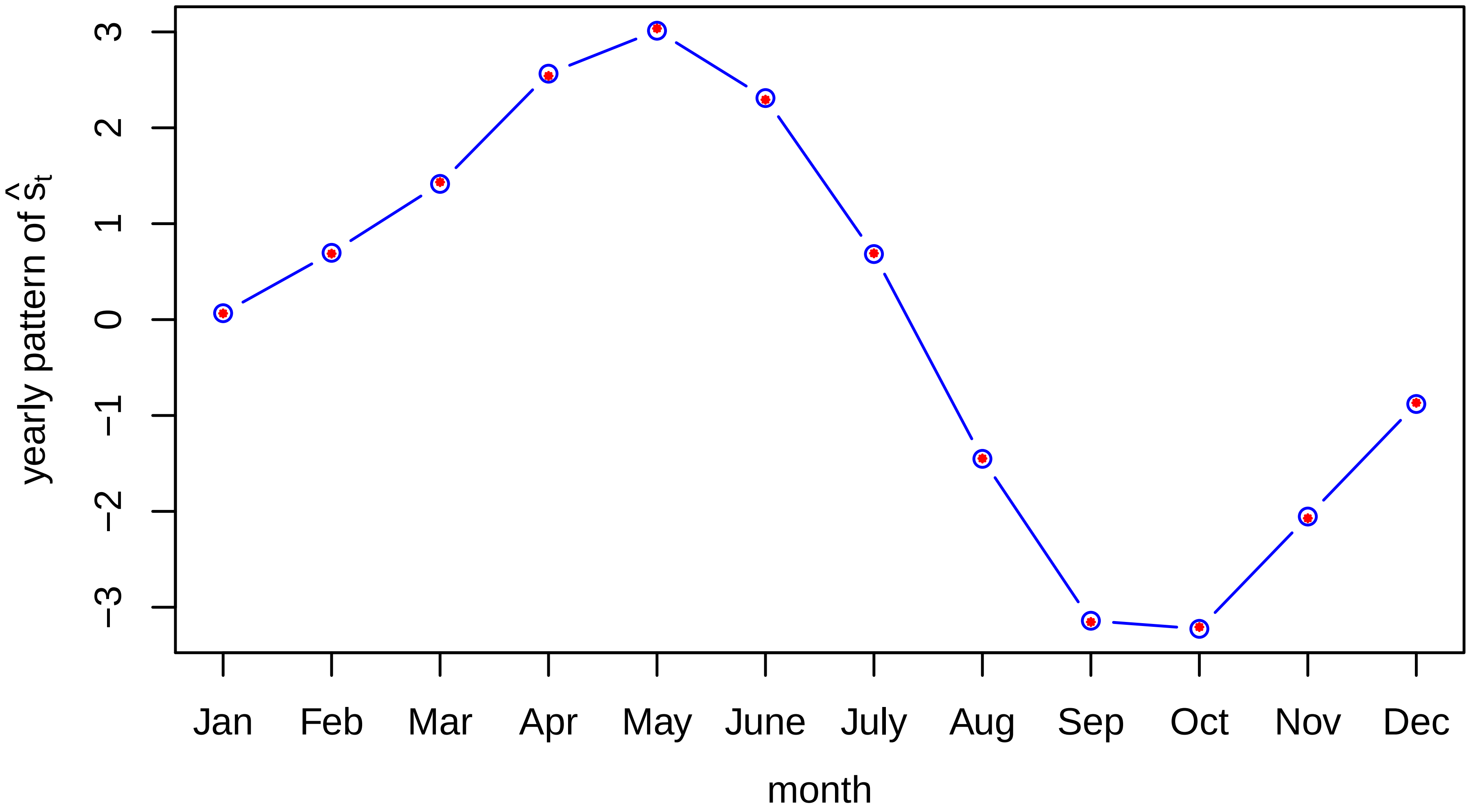
$J = 2$ Approximation to Seasonal Component



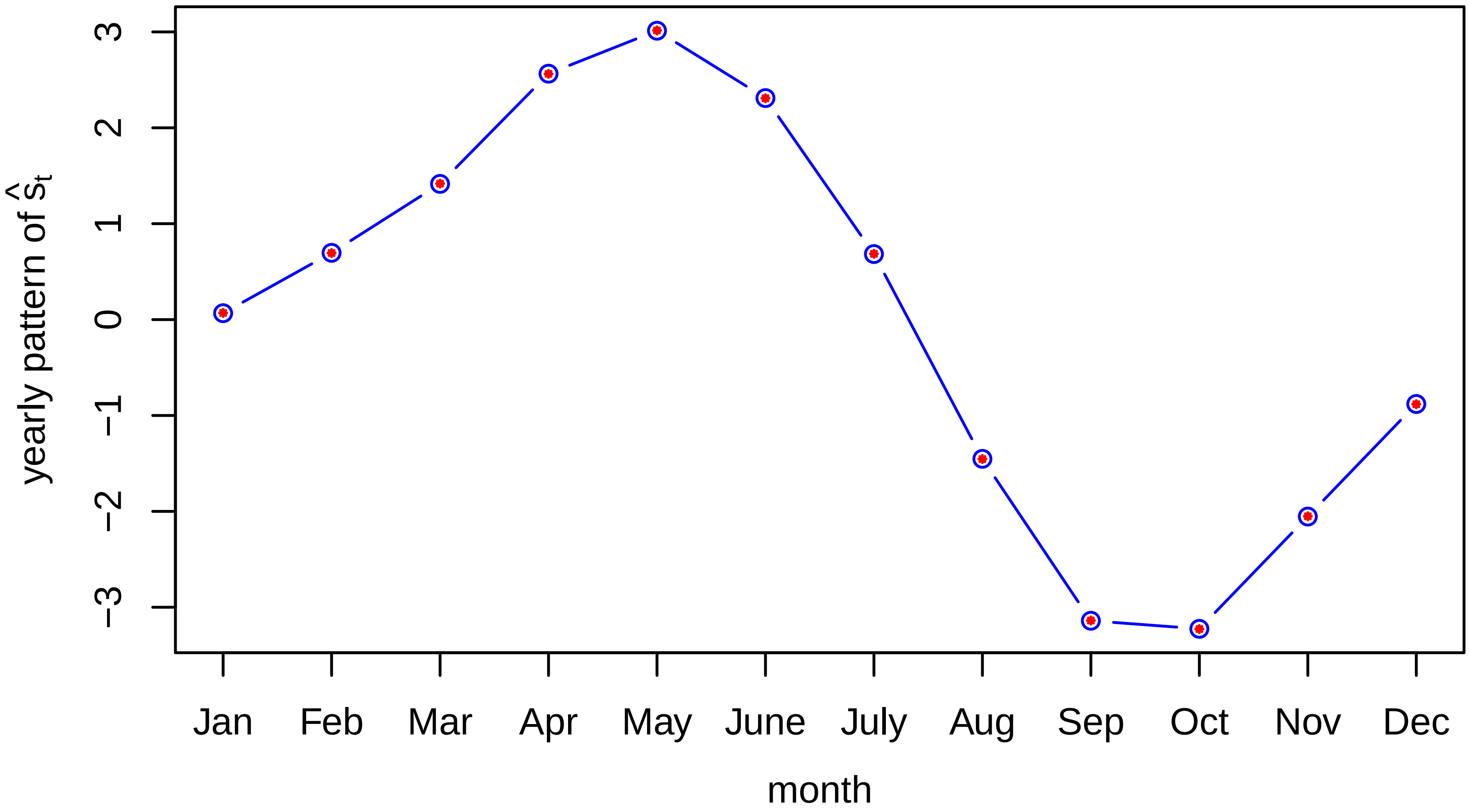
$J = 3$ Approximation to Seasonal Component



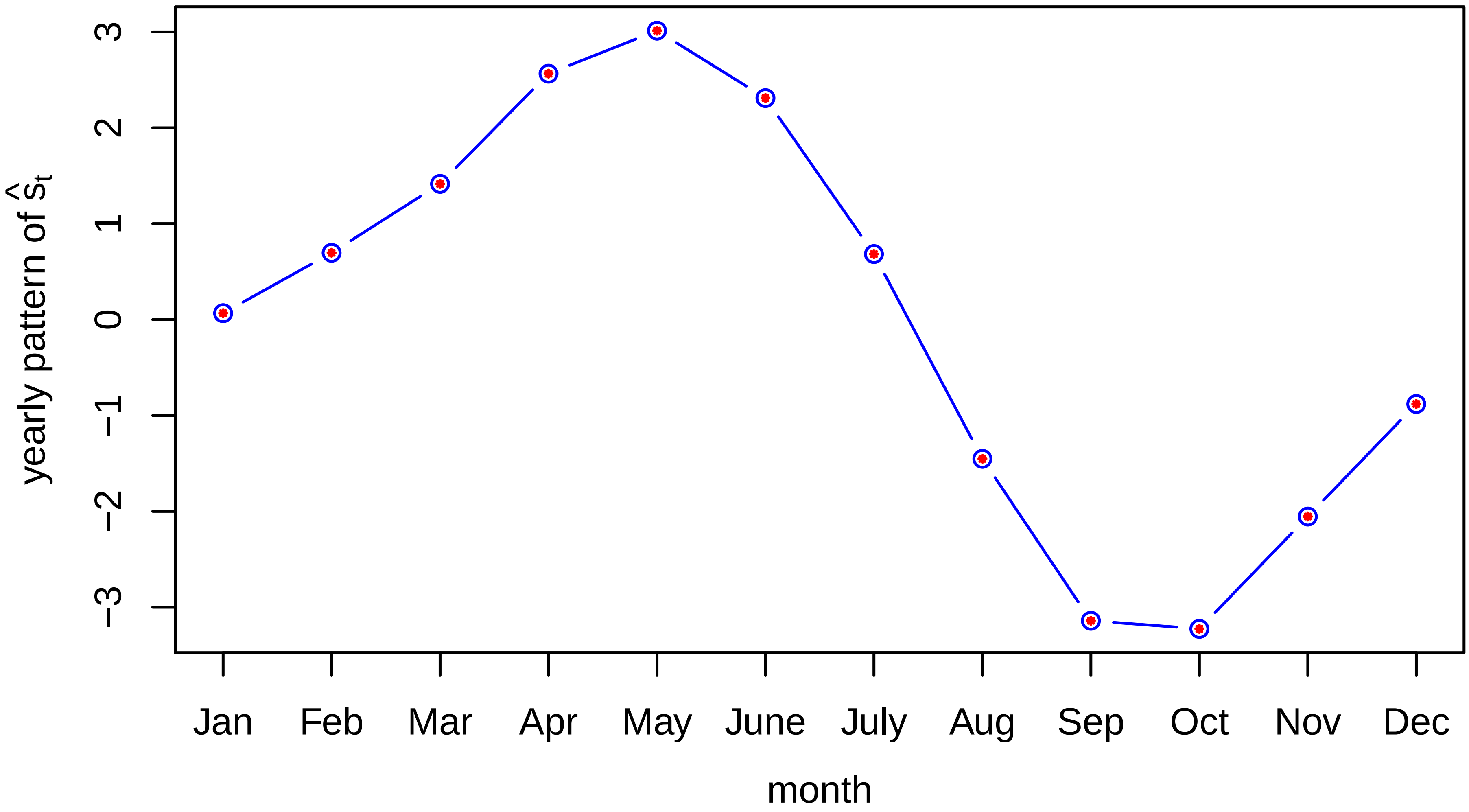
$J = 4$ Approximation to Seasonal Component



$J = 5$ Approximation to Seasonal Component



$J = 6$ Perfect Fit to Seasonal Component



Harmonic Regression and CO_2 Series: IV

- $J = 2$ approximation looks reasonable, so will entertain model

$$\begin{aligned} Y_t &= m_t + s_t + W_t \\ &= a + bt + ct^2 + \sum_{j=1}^{2} A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t) + W_t \end{aligned}$$

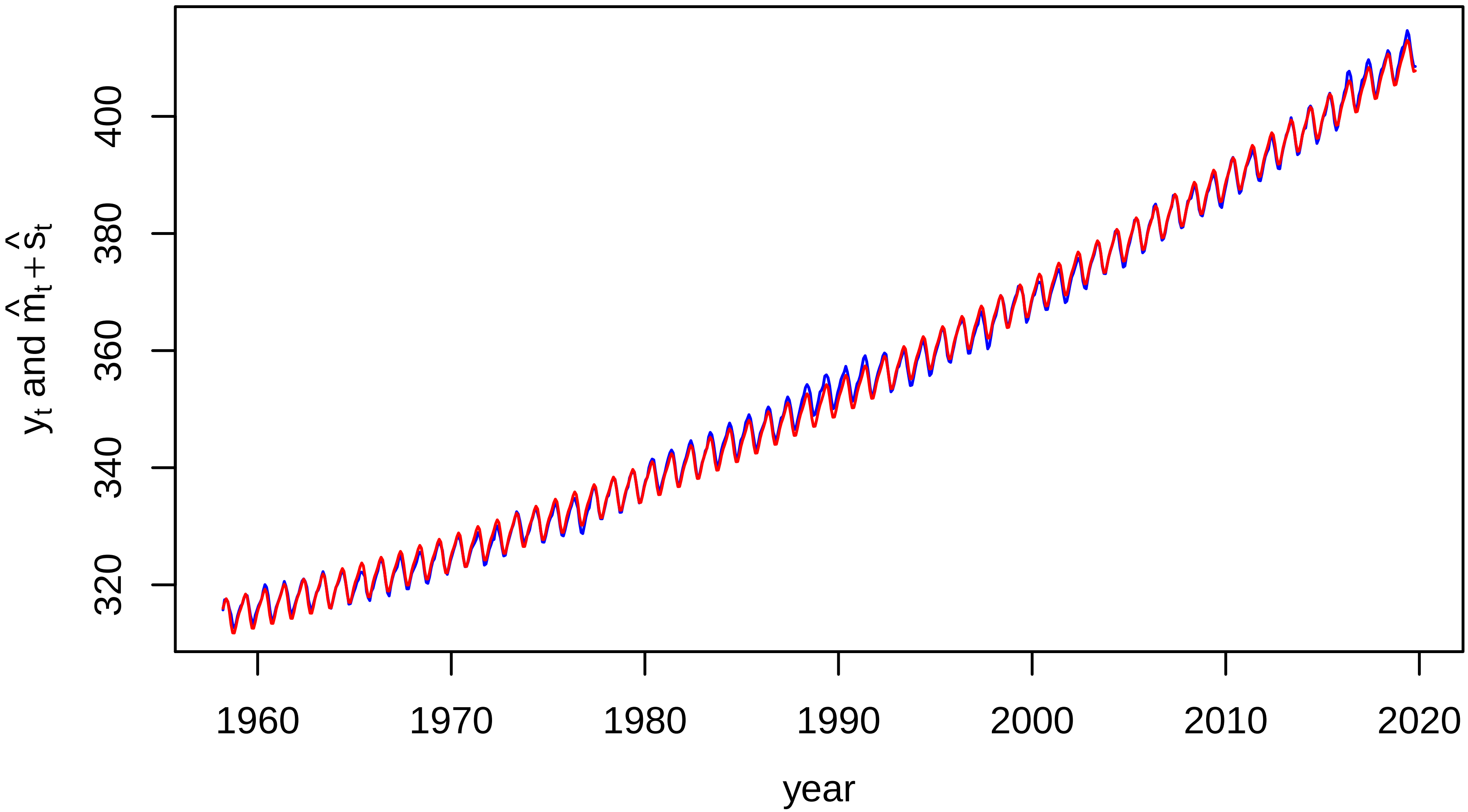
(A_j's & B_j's are coefficients for so-called harmonic regressors)

- start analysis by fitting model using OLS to get estimates

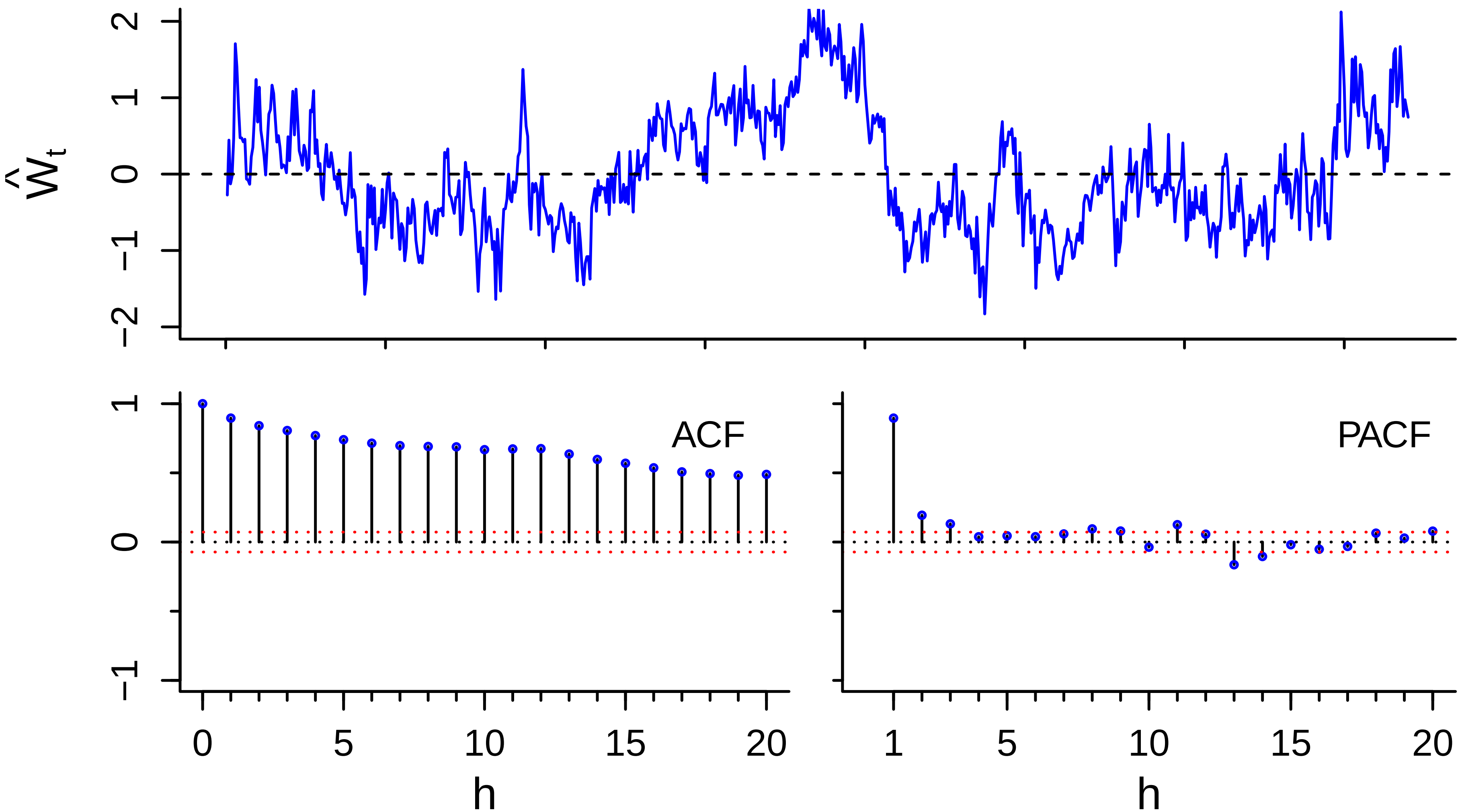
$$\hat{m}_t = \hat{a} + \hat{b}t + \hat{c}t^2 \text{ and } \hat{s}_t = \sum_{j=1}^{2} \hat{A}_j \cos(2\pi f_j t) + \hat{B}_j \sin(2\pi f_j t)$$

- following plots look at
  - $\hat{m}_t + \hat{s}_t$ as compared to Y_t
  - $\widehat{W}_t = Y_t - \hat{m}_t - \hat{s}_t$ (surrogates for W_t) and also $\nabla \widehat{W}_t$

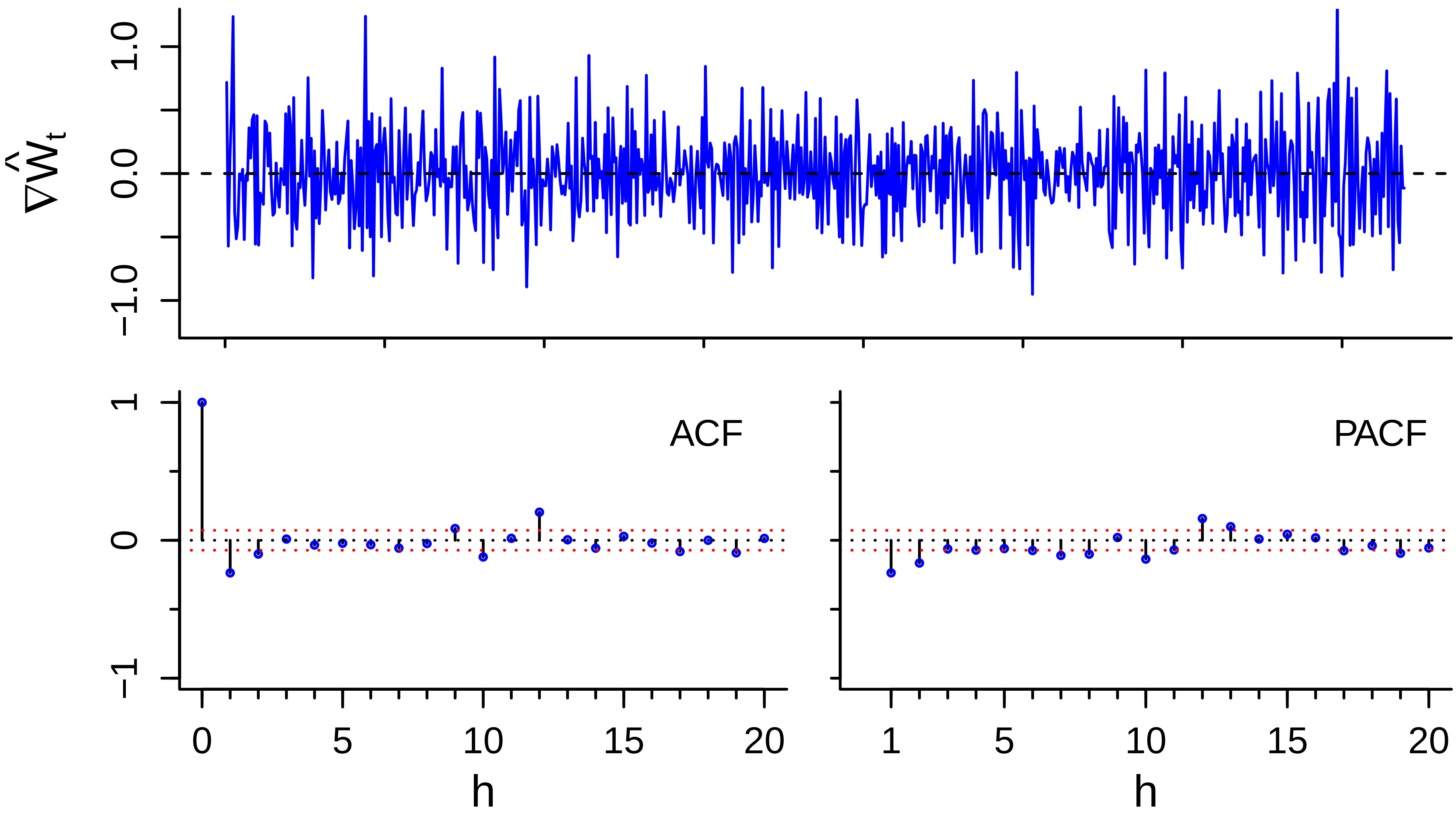
Parametric Estimate $\hat{m}_t + \hat{s}_t$ and CO_2 Series



Residuals $\widehat{W}_t$ from Parametric Fit of $m_t + s_t$



First Difference of Residuals $\widehat{W_t}$



Harmonic Regression and CO_2 Series: V

- sample ACFs & PACFs for $\widetilde{W}_t$ & $\nabla\widetilde{W}_t$ do not point to obvious simple model – more work needed to find reasonable model
- starting with $\widetilde{W}_t$, consideration of
  – AR(p), with $p = 1, \ldots, 35$ (`ar.mle` bombs for $p = 28$, 29, 30 & 33)
  – ARMA(p,q), with $p + q$ small
  – ARFIMA(p,δ,q), with $p + q$ small

  using

  – maximum likelihood to fit each model
  – AICC to evaluate individual models

  leads to somewhat unsatisfying AR(26) model

  – for details, see `R` code for this overhead

Harmonic Regression and CO_2 Series: VI

- ADF unit root test on $\widetilde{W}_t$ suggests need to difference, but not strongly so (p-value hovers around 0.05, but depends on choice of AR order p to be used with test)
- need for differencing also hinted at by δ estimates, which are close to upper limit of 1/2
- sample ACF & PACF for $\nabla \widetilde{W}_t$ have large values at lag $h = 12$, suggesting that deterministic s_t might be too simplistic
- SARIMA model with either $d = 1$ or $D = 1$ (or both) worthy of consideration (more work is needed!)
- bottom line: finding suitable model for W_t for atomic clock data relatively easy, but finding one for CO_2 series more of a challenge