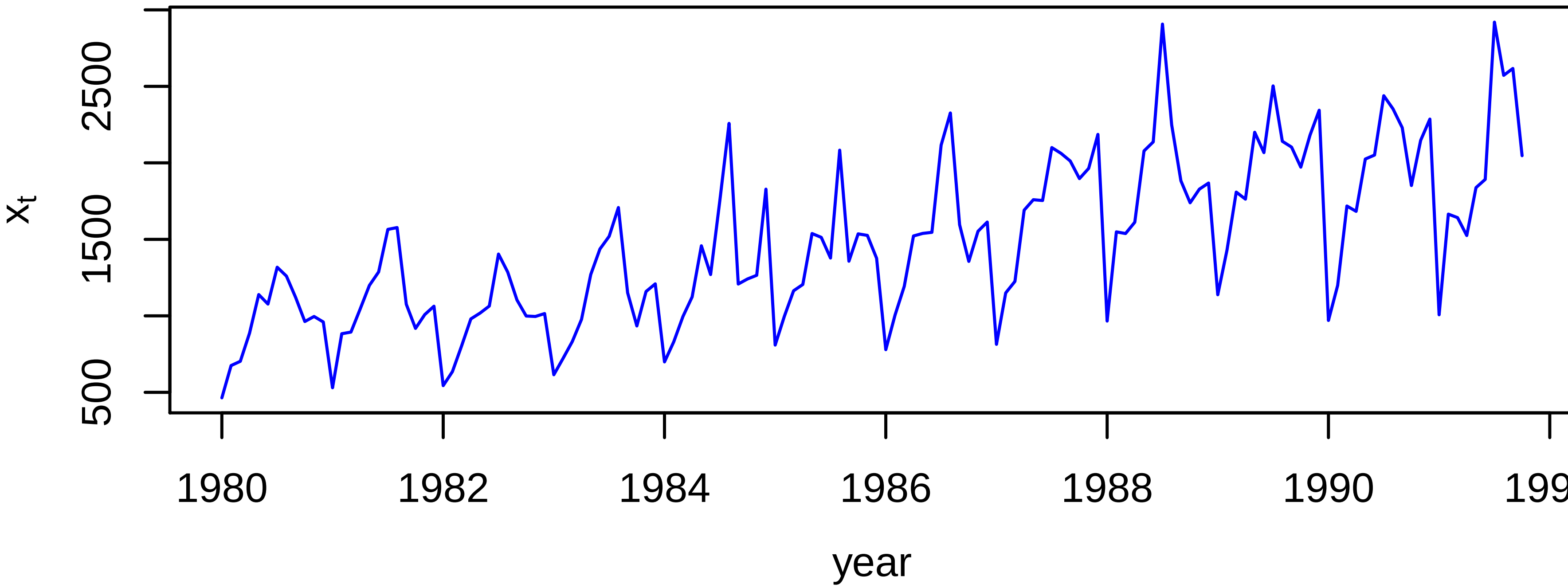
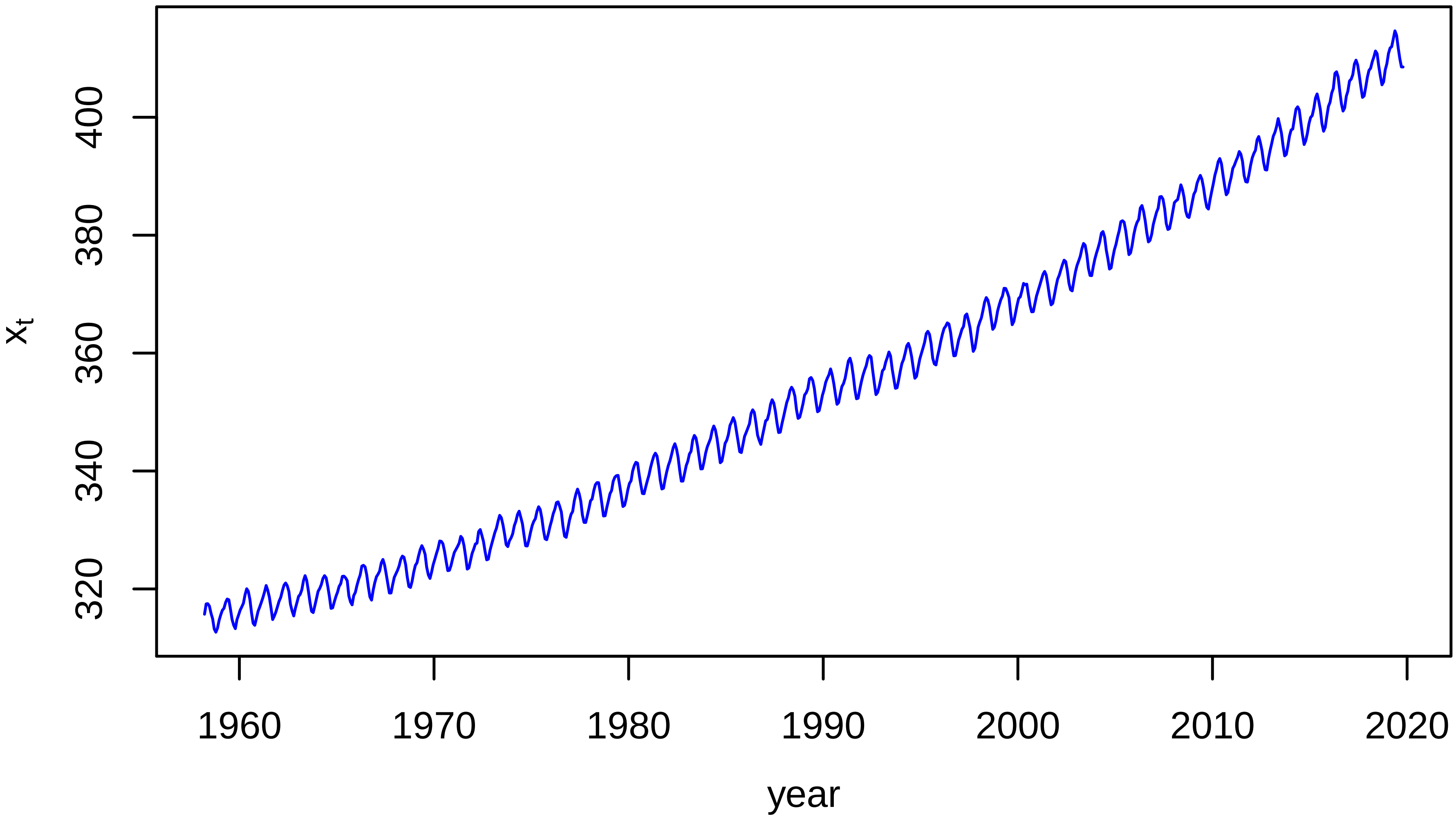
Models with Trend and Seasonality: I

- time series often exhibit trends & seasonal variations, for which a stationary model might be inappropriate
- example is Australian monthly red wine sales



2nd Example: CO_2 Series from Mauna Loa, Hawaii



Models with Trend

- as a first step toward modeling time series with trends, consider

$$X_t = m_t + Y_t,$$

where $\{m_t\}$ is a slowly varying (smooth) sequence (the trend component), while $\{Y_t\}$ is a stationary process with mean zero

- if $\{m_t\}$ is deterministic, then

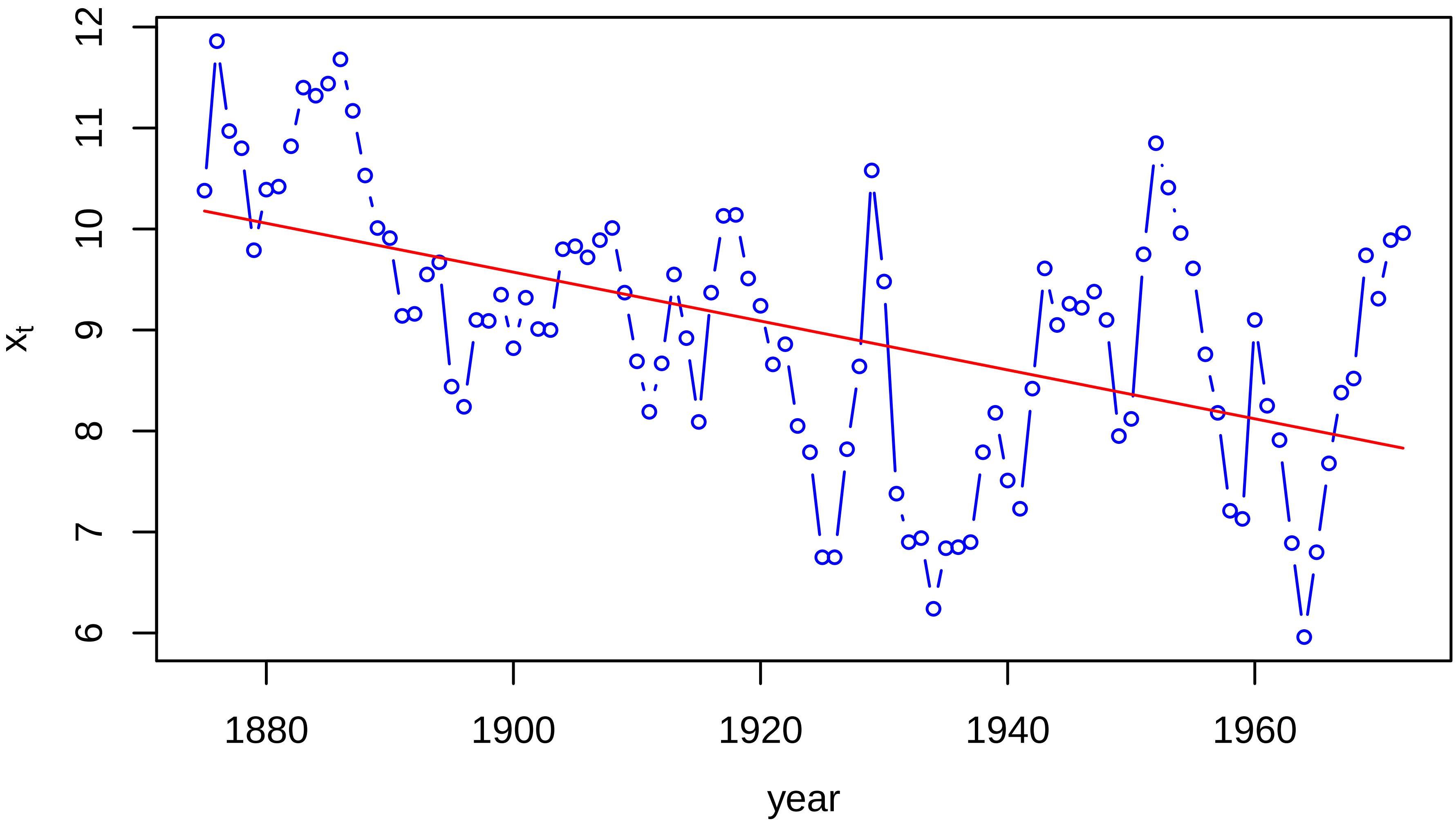
$$E\{X_t\} = E\{m_t\} + E\{Y_t\} = m_t$$

- one popular specification for $\{m_t\}$ is a low-order polynomial, e.g.,

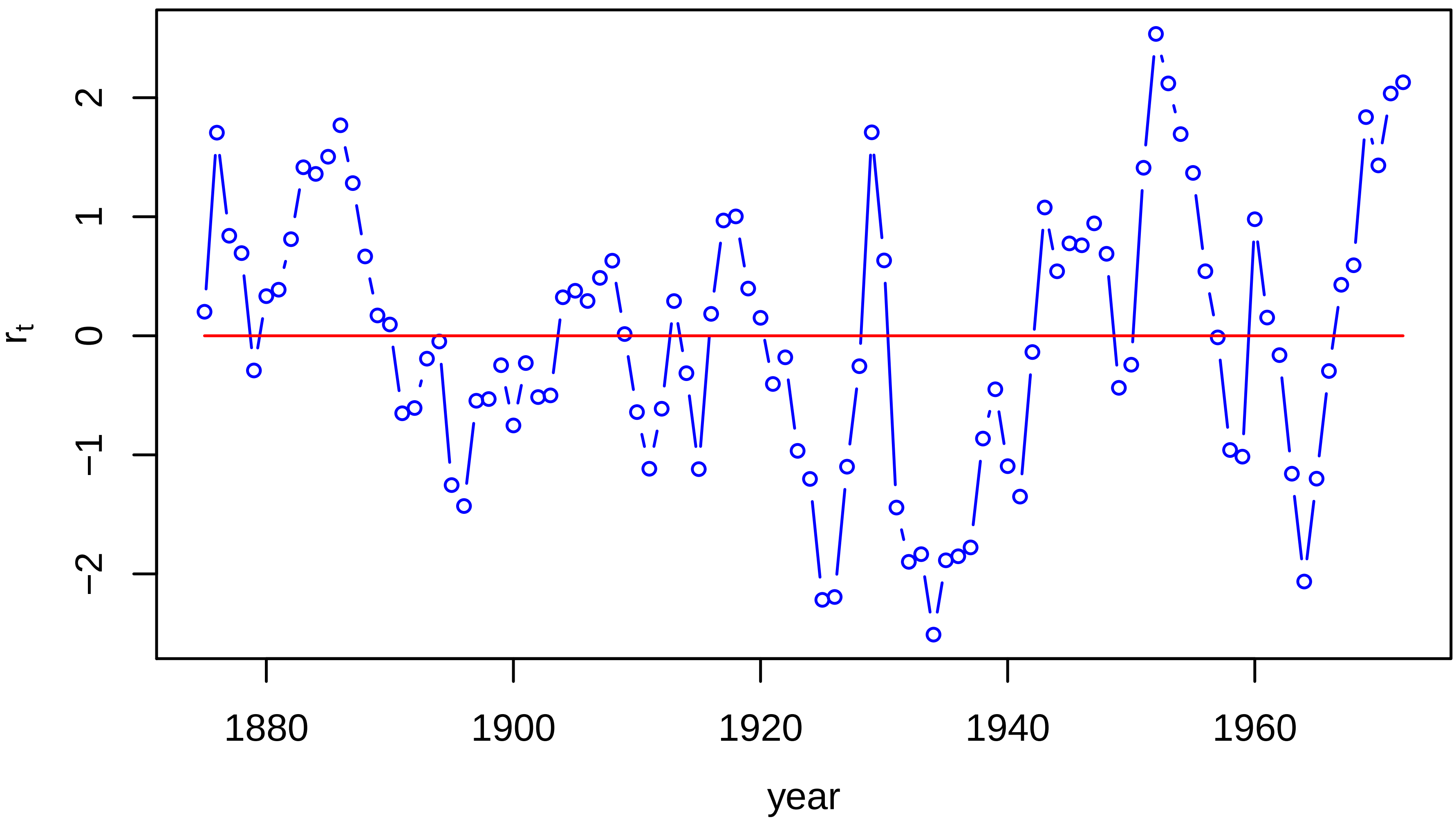
$$\begin{aligned} m_t &= c_0 + c_1 t && \text{(linear)} \\ m_t &= c_0 + c_1 t + c_2 t^2 && \text{(quadratic)} \\ m_t &= c_0 + c_1 t + c_2 t^2 + c_3 t^3 && \text{(cubic)} \end{aligned}$$

- can estimate c_j's via least squares: minimize $\sum_t (x_t - m_t)^2$

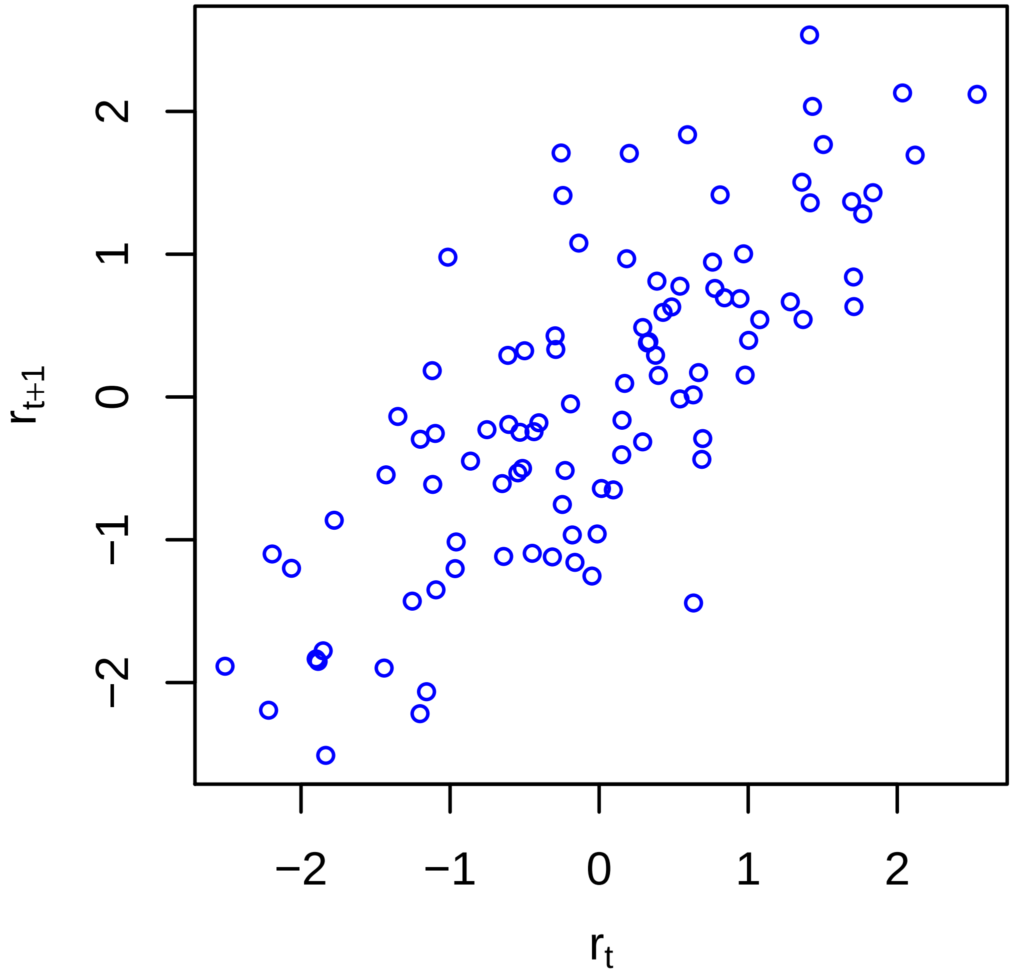
Level of Lake Huron (1875–1972)



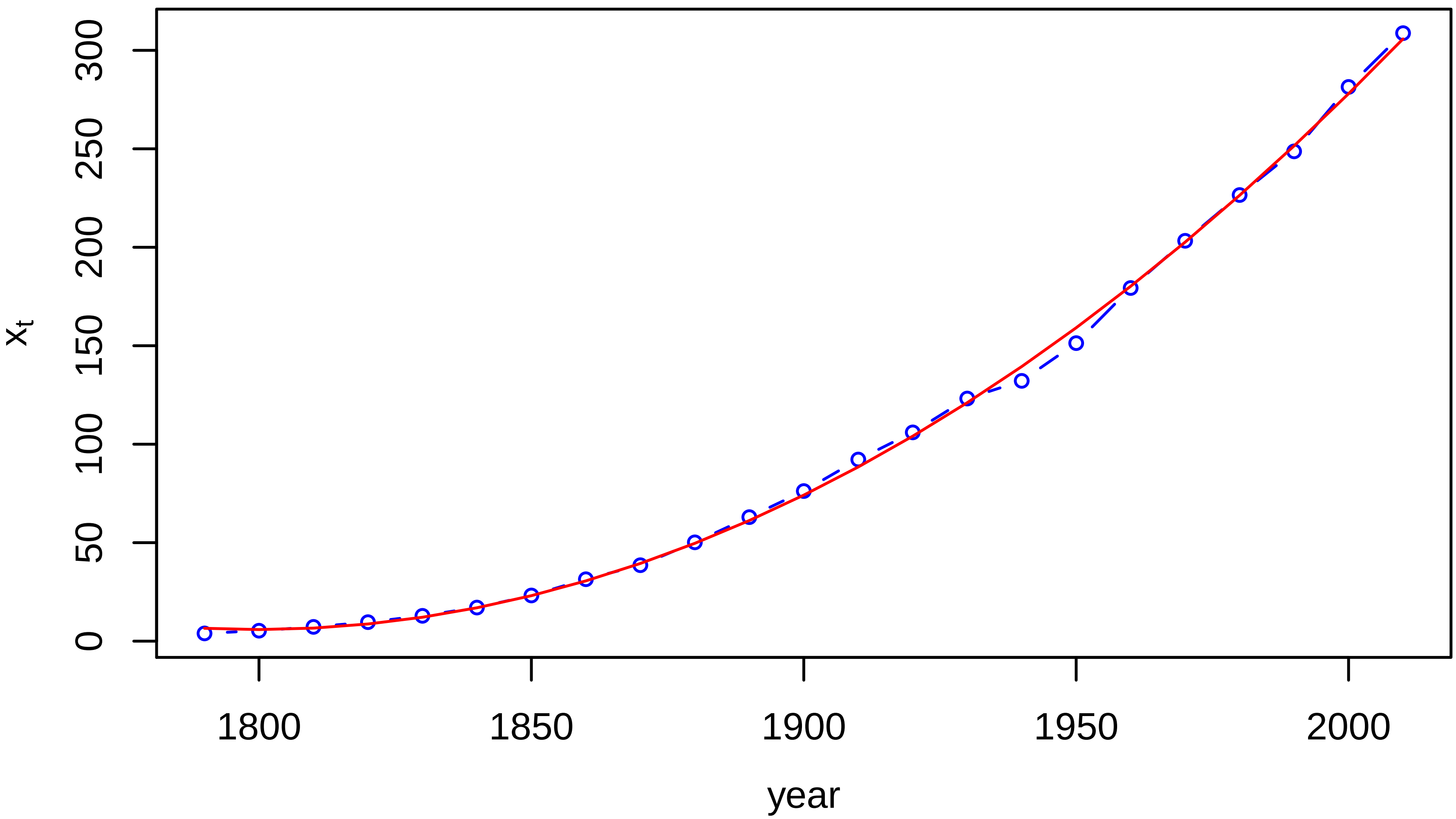
Residuals $r_t = x_t - \hat{c}_0 - \hat{c}_1 t$ from Least Squares Fit



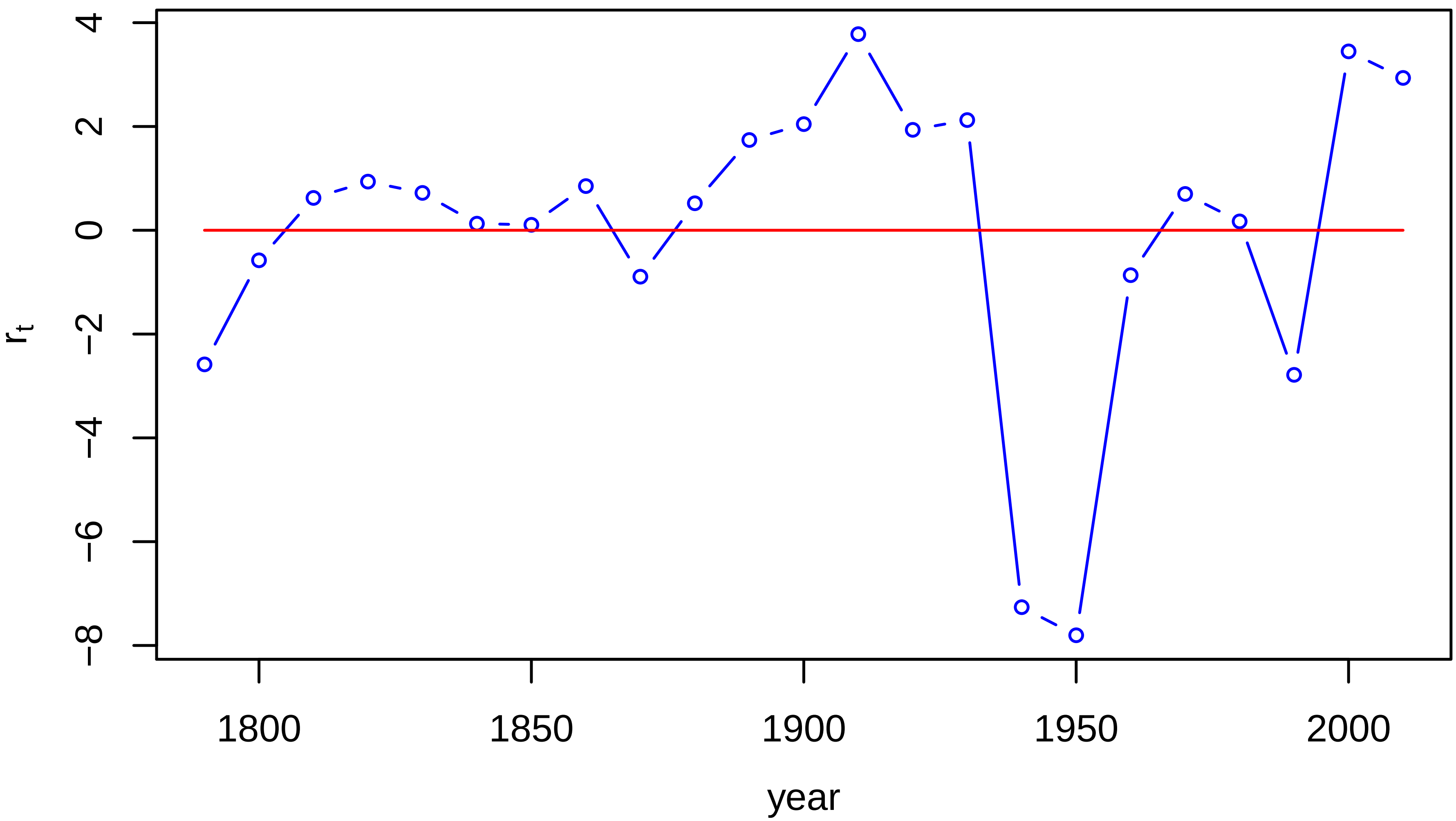
Unit Lag Scatter Plot of Residuals $\{r_t\}$



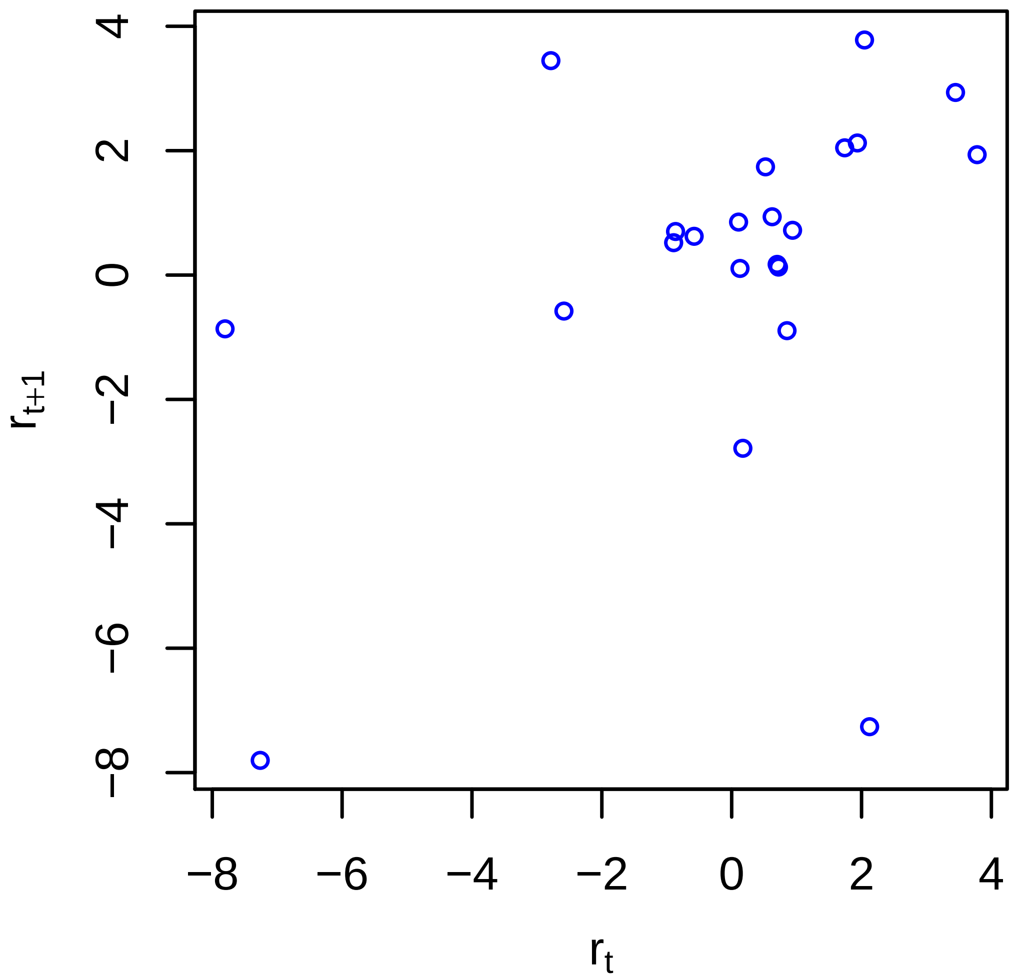
Population of USA in Millions (1790–2010)



Residuals $r_t = x_t - \hat{c}_0 - \hat{c}_1 t - \hat{c}_2 t^2$ from LS Fit



Unit Lag Scatter Plot of Residuals $\{r_t\}$



Models with Trend and Seasonality: II

- to handle a time series with a trend and a seasonal component, can entertain model

$$X_t = m_t + s_t + Y_t$$

  - $\{m_t\}$ is the trend ($m_t = \mu$ is OK, i.e., a degenerate trend);
  - $\{s_t\}$ is seasonal component with known period d (i.e., $s_{t+d} = s_t$ for all t) satisfying

$$\sum_{j=1}^{d} s_{t+j} = 0 \text{ for all } t\text{; and}$$

  - $\{Y_t\}$ is a stationary process with mean zero
- assuming $\{m_t\}$ & $\{s_t\}$ to be deterministic, we have

$$E\{X_t\} = E\{m_t\} + E\{s_t\} + E\{Y_t\} = m_t + s_t$$

General Approach to Simple Time Series Modeling

- plot x_t and check for presence of
  - trend and seasonal component
  - sharp changes in behaviour (model subseries individually?)
  - outliers (take these out somehow?)
- if trend & seasonal component present, entertain model

$$X_t = m_t + s_t + Y_t$$

  and estimate m_t & s_t somehow (denote estimates by $\hat{m}_t$ & $\hat{s}_t$)
- create residuals $r_t \stackrel{\text{def}}{=} x_t - \hat{m}_t - \hat{s}_t$ (surrogate for Y_t)
- determine model for residuals somehow
- $\hat{m}_t$, $\hat{s}_t$ and residual model can be used for, e.g., forecasting
- note: might need to transform x_t to get approach to work

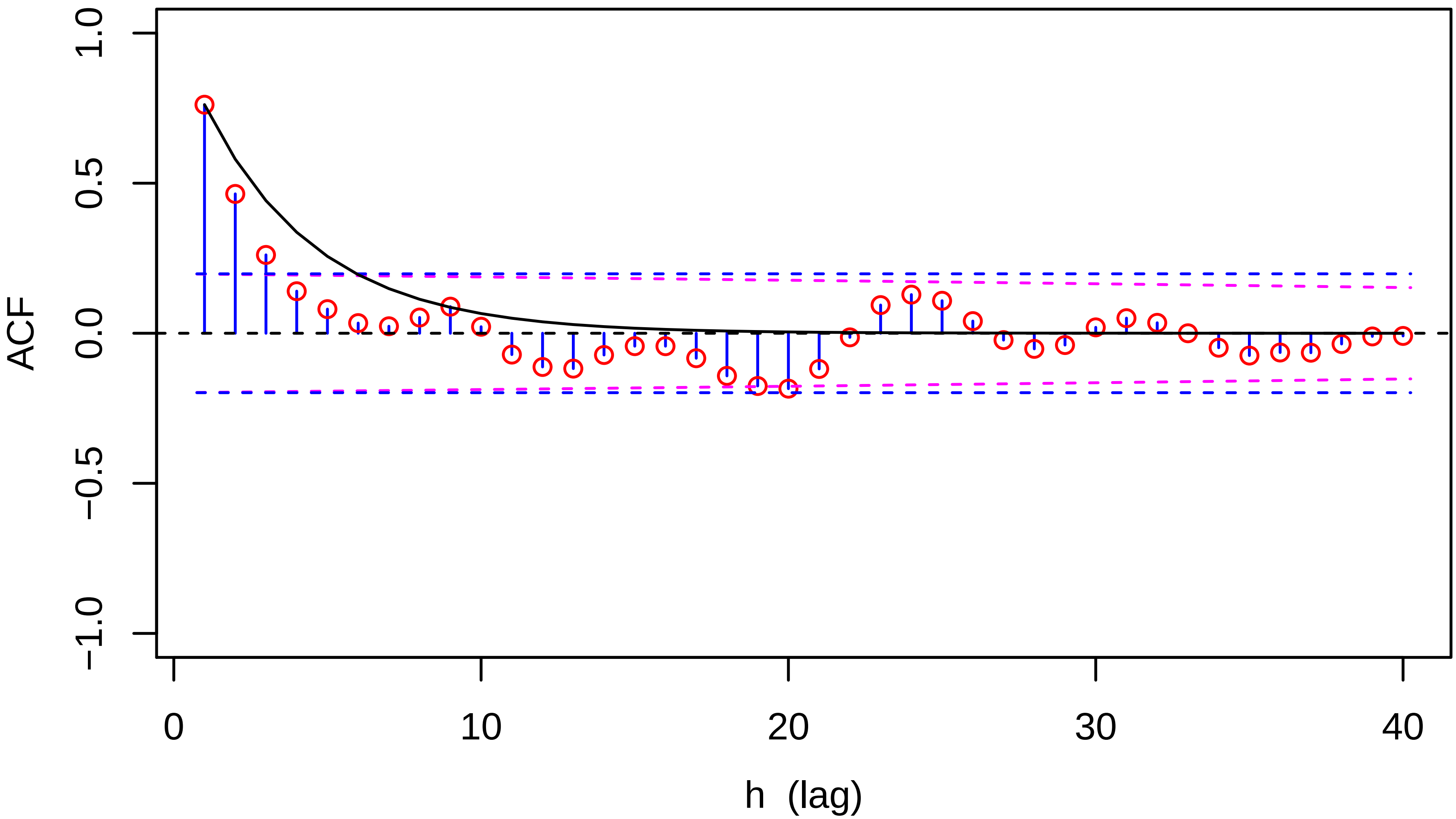
Model for Lake Huron Levels: I

- recall our preliminary assessment of Lake Huron levels in terms of the model

$$X_t = m_t + Y_t = c_0 + c_1 t + Y_t$$

- residuals $r_t = x_t - \hat{c}_0 - \hat{c}_1 t$ after detrending can be regarded as surrogates for Y_t's, but unit-lag scatter plot suggests r_t's are not consistent with hypothesis that $\{Y_t\}$ is IID noise
- let's now look at sample ACF for r_t's

Sample ACF for Residuals $\{r_t\}$



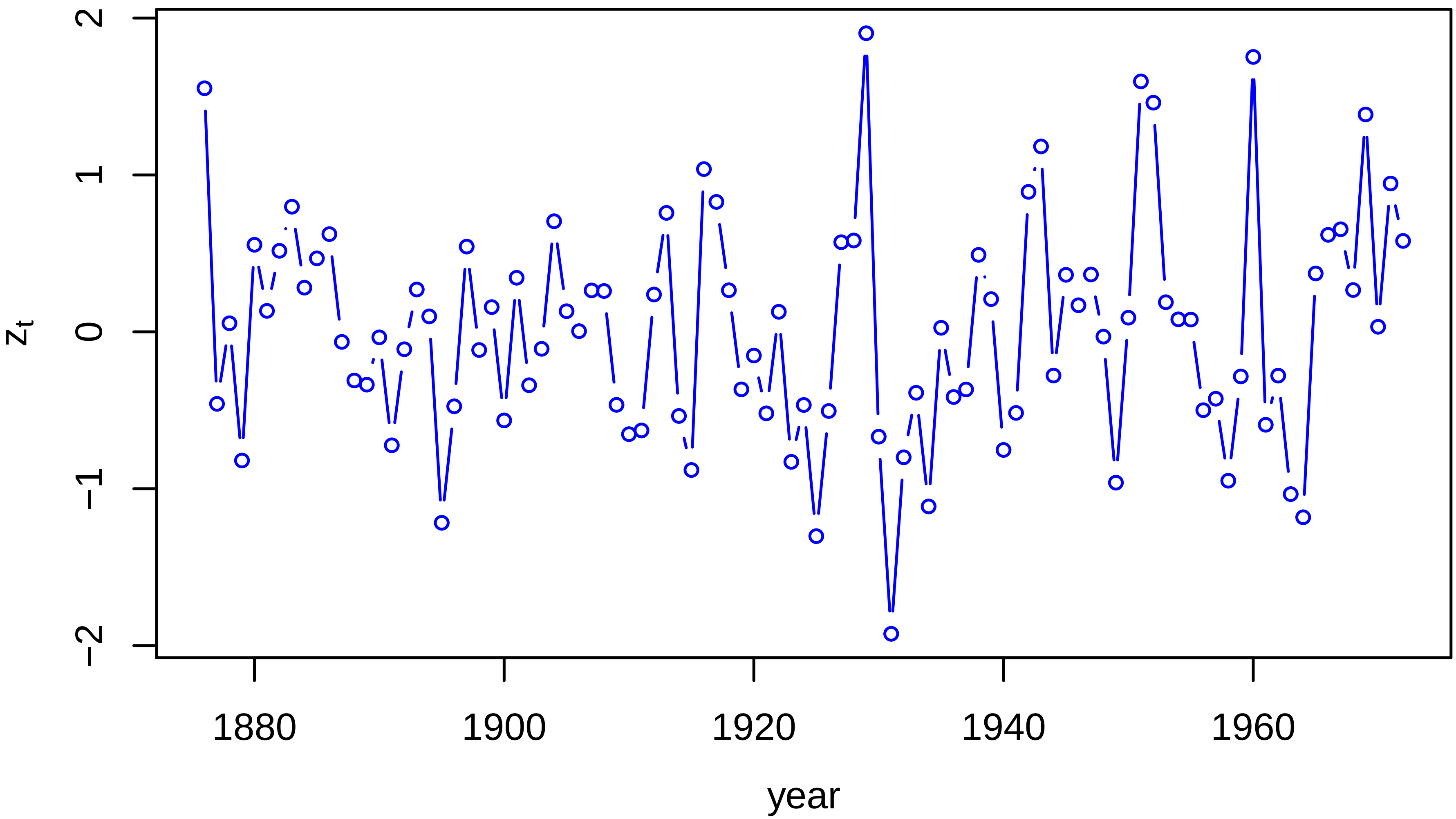
Model for Lake Huron Levels: II

- sample ACF suggests IID noise hypothesis not viable (p-value for $\hat{\rho}_r(1)$ is 4.7×10^{-14})
- for AR(1) process, $\rho(h) = \phi^{|h|}$, so, if we want to entertain this model for the r_t's, can estimate ϕ using $\hat{\phi} \stackrel{\text{def}}{=} \hat{\rho}_r(1) \doteq 0.762$
- black curve on previous overhead shows $\hat{\phi}^h$ versus h – agreement with $\hat{\rho}_r(h)$ for $h \geq 2$ not perfect, but perhaps not unreasonable when sampling variability is taken into account
- will thus entertain model

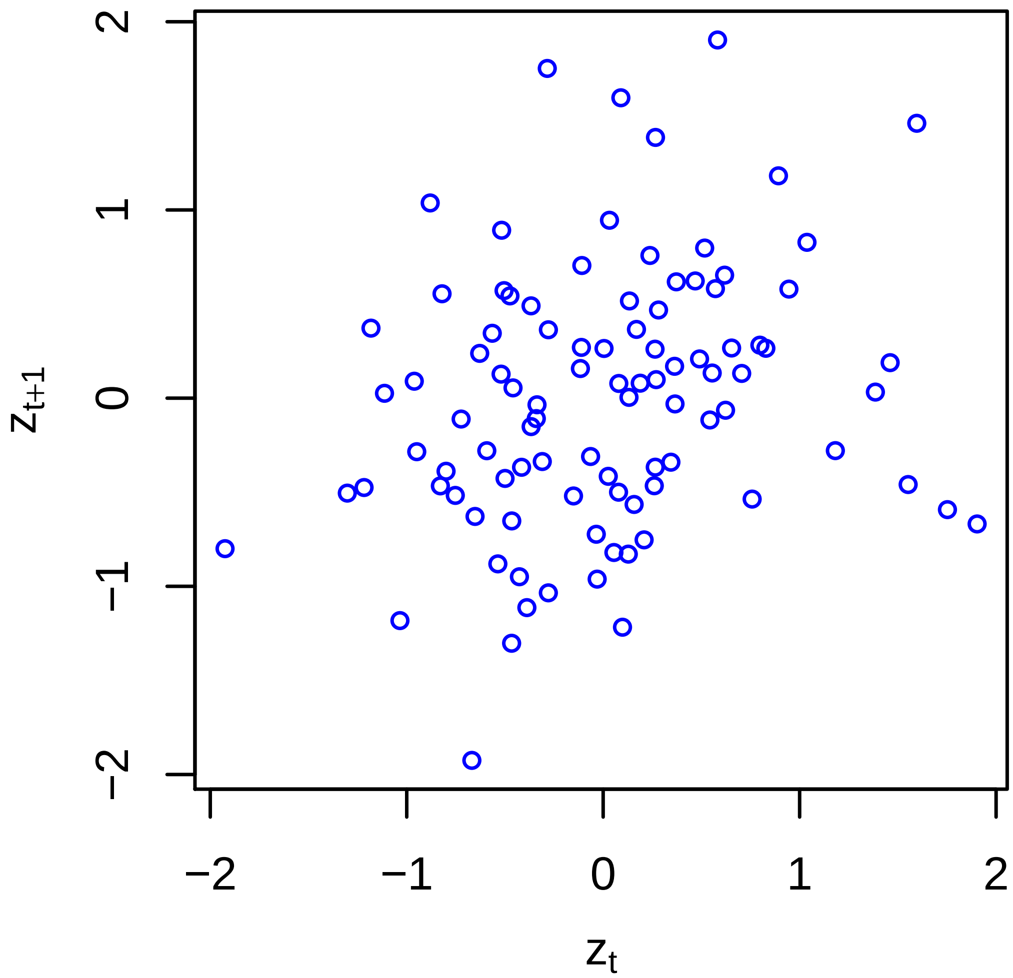
$$R_t = \phi R_{t-1} + Z_t$$

  for $\{r_t\}$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- if AR(1) model is viable, then $z_t = r_t - \hat{\phi} r_{t-1}$, $t = 2, 3, \ldots, n$, should resemble a white noise process

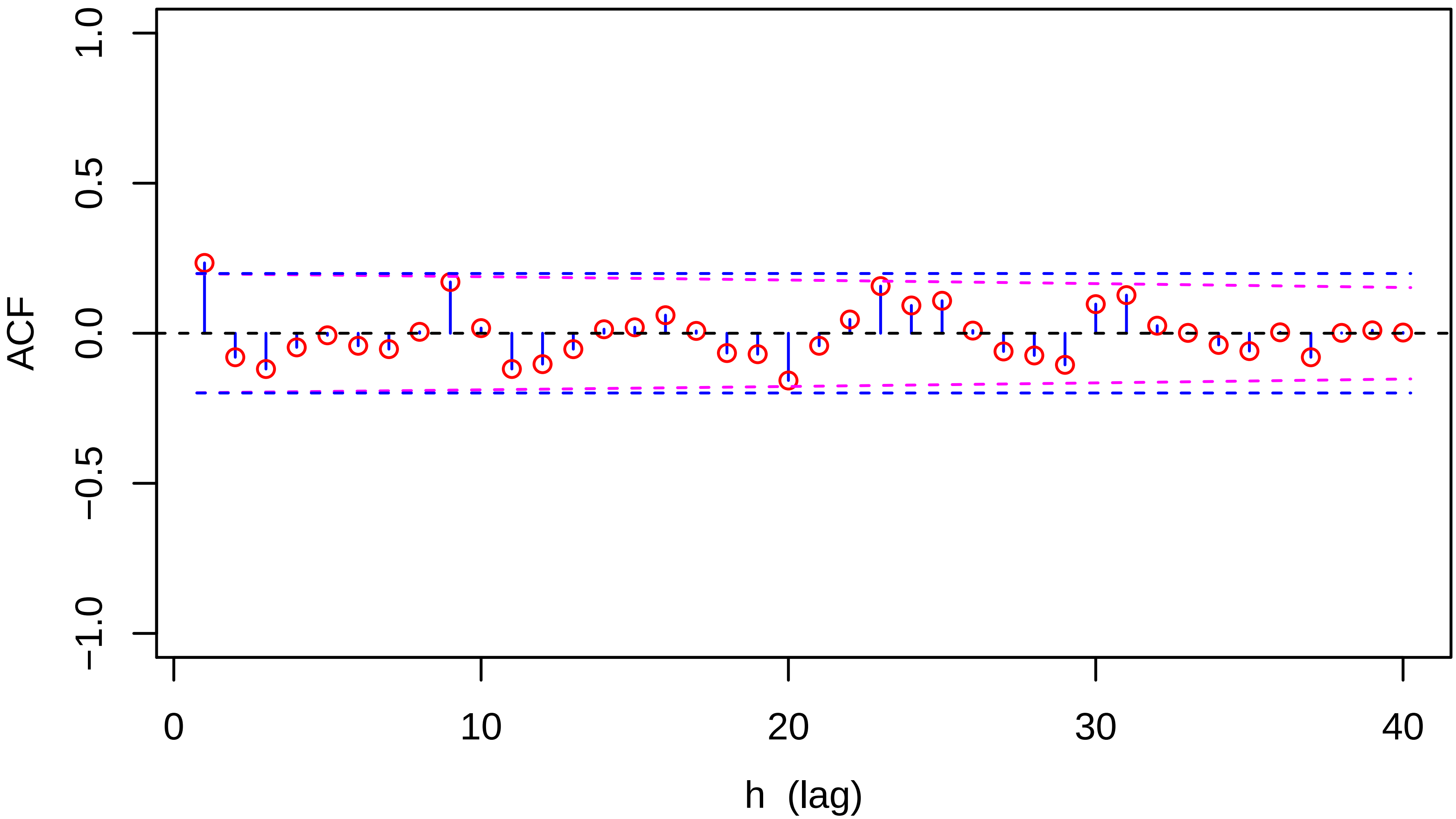
AR(1) Residuals $z_t = r_t - \hat{\phi} r_{t-1}$



Unit Lag Scatter Plot of AR(1) Residuals $\{z_t\}$



Sample ACF for AR(1) Residuals $\{z_t\}$

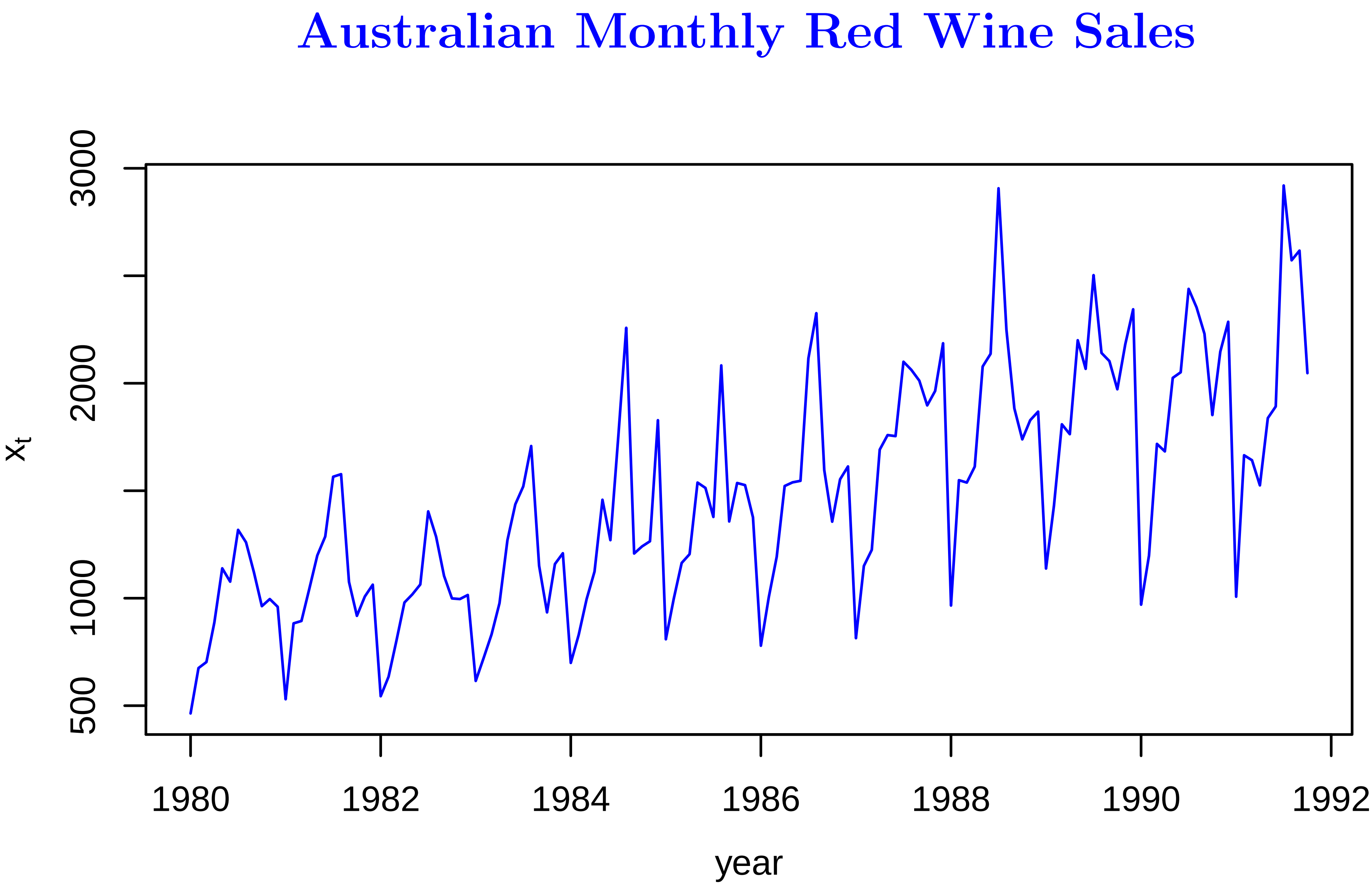


Model for Lake Huron Levels: III

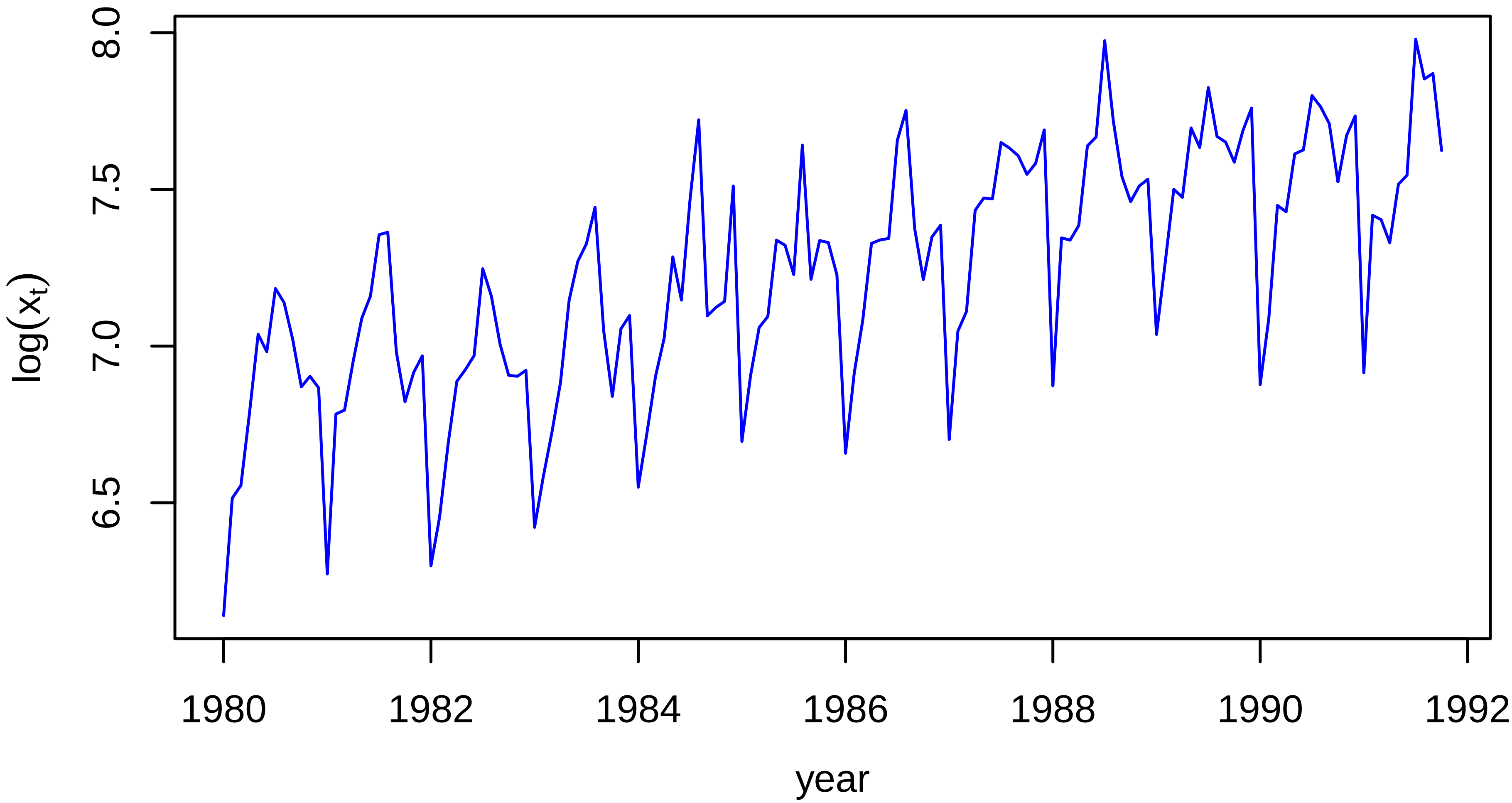
- three comments
  - only $\hat{\rho}_z(1)$ outside of $\pm 1.96/\sqrt{n}$ bounds for IID noise (its p-value is 0.02 and hence only mildly worrisome)
  - will reconsider null hypothesis that z_t's are realization of IID process using a battery of statistical tests discussed later on
  - Brockwell & Davis suggest that a better fit for $\{r_t\}$ is a second-order autoregressive process (more on this model later)

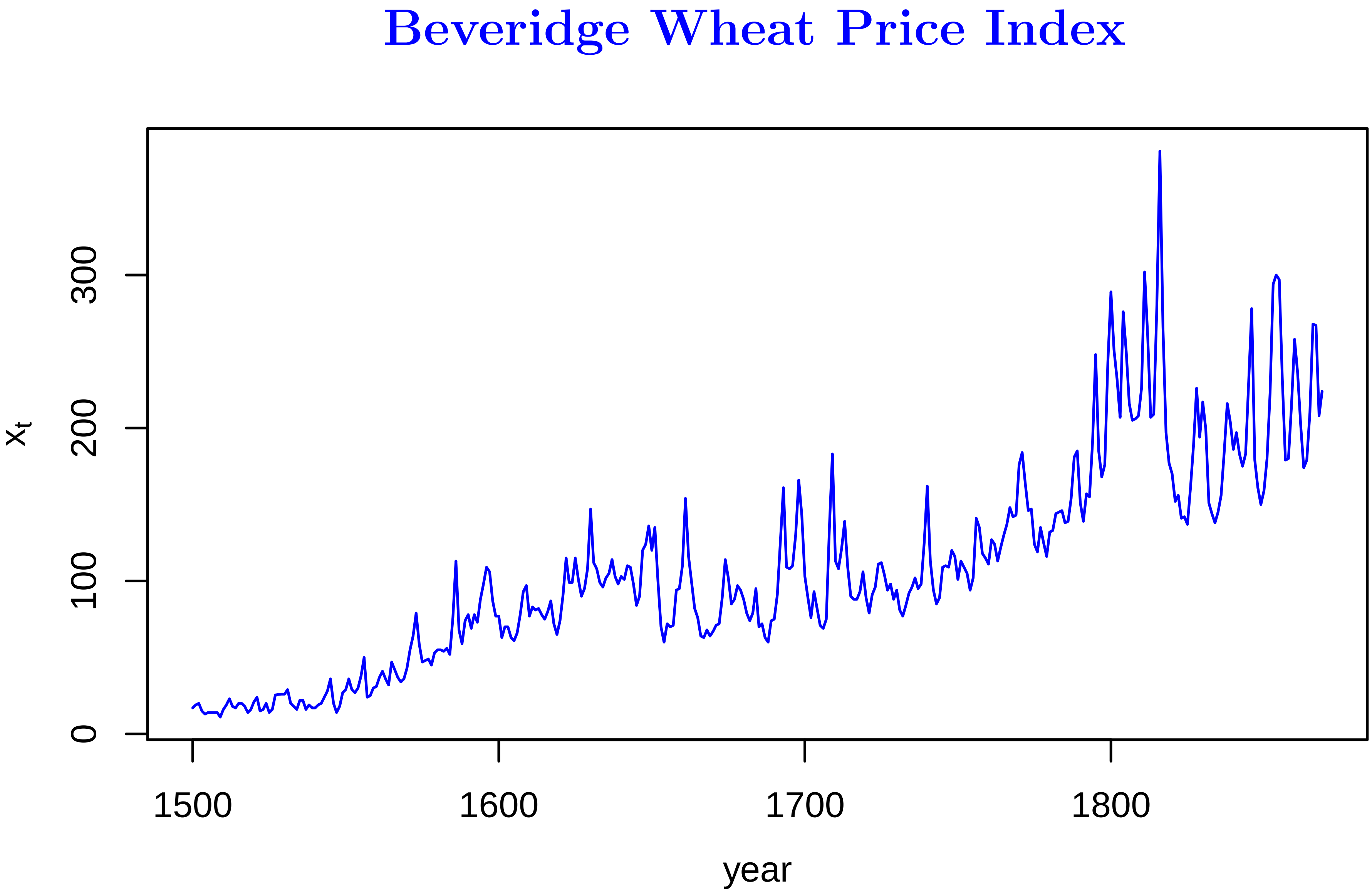
Classical Decomposition Model

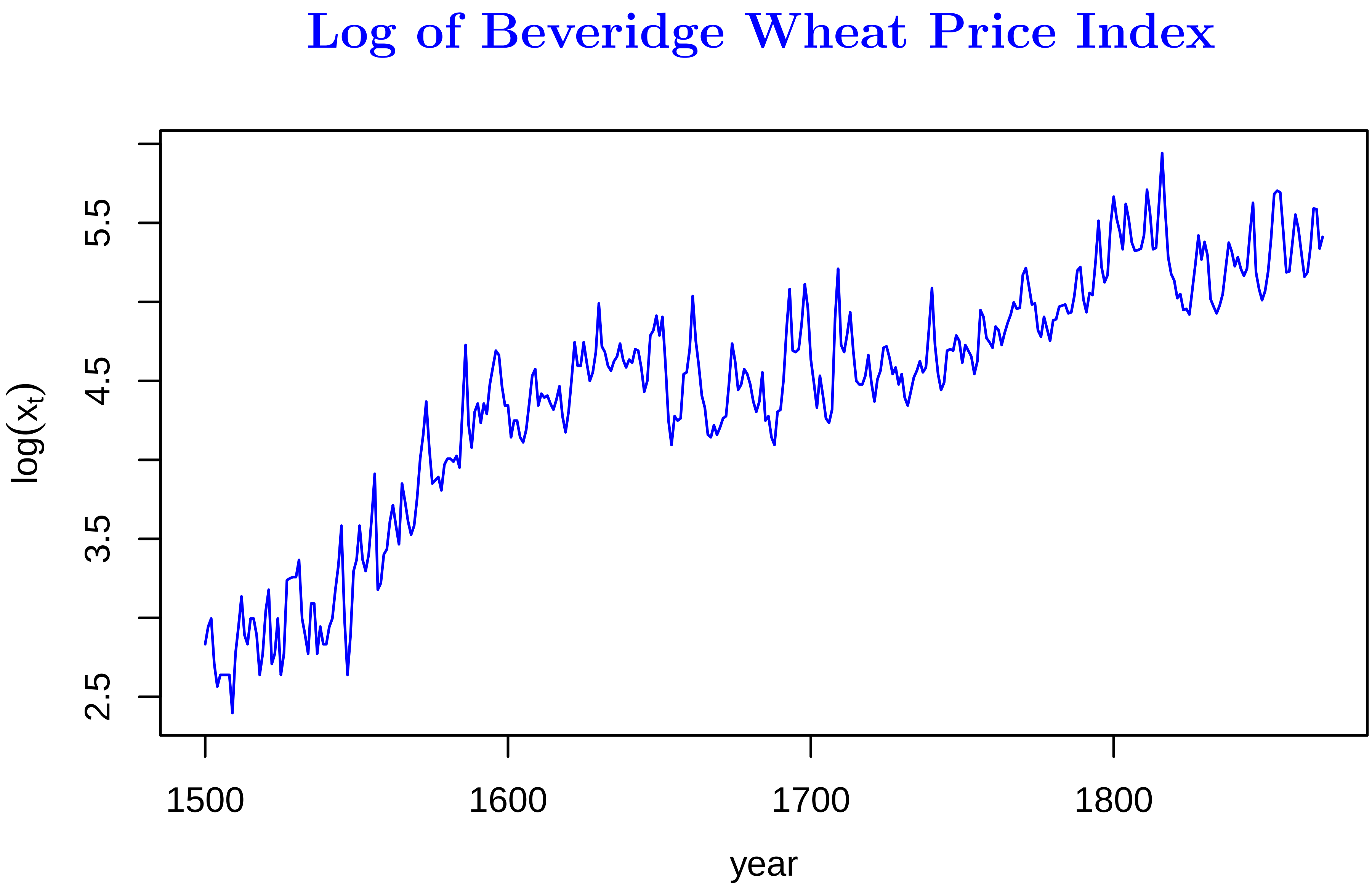
- consider time series $\{x_t\}$ for which classical decomposition model
$$X_t = m_t + s_t + Y_t$$
might be appropriate, where
  - $\{m_t\}$ is trend;
  - $\{s_t\}$ is periodic with known period d (i.e., $s_{t+d} = s_t$ for all $t \in \mathbb{Z}$); and
  - $\{Y_t\}$ is a mean-zero stationary process
- some time series can be handled under this model after application of an appropriate transform, e.g., $\log(x_t)$
- two examples
  - Australian monthly red wine sales
  - Beveridge wheat price index (yearly from 1500 to 1869)


Australian Monthly Red Wine Sales
3000
2000
1000
500
x_t
1980
1982
1984
1986
1988
1990
1992
year

Log of Australian Monthly Red Wine Sales




Beveridge Wheat Price Index
x_t
0
100
200
300
1500
1600
1700
1800
year


Log of Beveridge Wheat Price Index
log(x_t)
2.5
3.5
4.5
5.5
1500
1600
1700
1800
year

Trend & Seasonal Estimation and Elimination: I

- one approach: estimate deterministic components via, say, $\hat{m}_t$ and $\hat{s}_t$, and use these to form residuals

$$r_t = x_t - \hat{m}_t - \hat{s}_t,$$

with the hope that $\{r_t\}$ can be considered to be a realization of a stationary process that is a surrogate for $\{Y_t\}$ in the model

$$X_t = m_t + s_t + Y_t$$

- second approach (Box & Jenkins): apply appropriate differencing operations to $\{x_t\}$ that in effect eliminate $\{m_t\}$ and $\{s_t\}$
- will now illustrate these two approaches (estimation and elimination), focusing first on the simpler model with trend, but no seasonal component:

$$X_t = m_t + Y_t$$

Trend Estimation via Two-Sided Filters: I

- consider a sequence $\{a_j : j = -q, \ldots, q\}$ of length $2q + 1$, where a_j's are real-valued, sum to unity and $a_{-j} = a_j$
- given time series $\{x_t : t = 1, \ldots, n\}$, use $\{a_j\}$ to create a new time series $\{w_t\}$ via

$$w_t = \sum_{j=-q}^{q} a_j x_{t-j}, \quad t = q+1, \ldots, n-q$$

- mapping from $\{x_t\}$ to $\{w_t\}$ is called a filter
  - $\{x_t\}$ is input to filter
  - $\{w_t\}$ is output from filter
  - $\{a_j\}$ represents the filter and is called the impulse response sequence in the engineering literature (there are other ways to represent a filter)

Trend Estimation via Two-Sided Filters: II

- let $a_j = 1/(2q+1)$ so that

$$w_t = \sum_{j=-q}^{q} a_j x_{t-j} = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}$$

- above defines a two-sided moving average filter
- under model $X_t = m_t + Y_t$, have $w_t \approx m_t$ if trend is approximately locally linear around t and if average of error terms about t is close to zero
- hence $\{w_t : t = q+1, \ldots, n-q\}$ is an estimate of $\{m_t : t = q+1, \ldots, n-q\}$, but estimates of m_1, …, m_q and m_{n-q+1}, …, m_n are also needed

Trend Estimation via Two-Sided Filters: III

- to estimate remaining $2q$ values, let input to filter be the following sequence of length $n + 2q$:

$$x_{-q+1}, \ldots, x_0, x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+q},$$

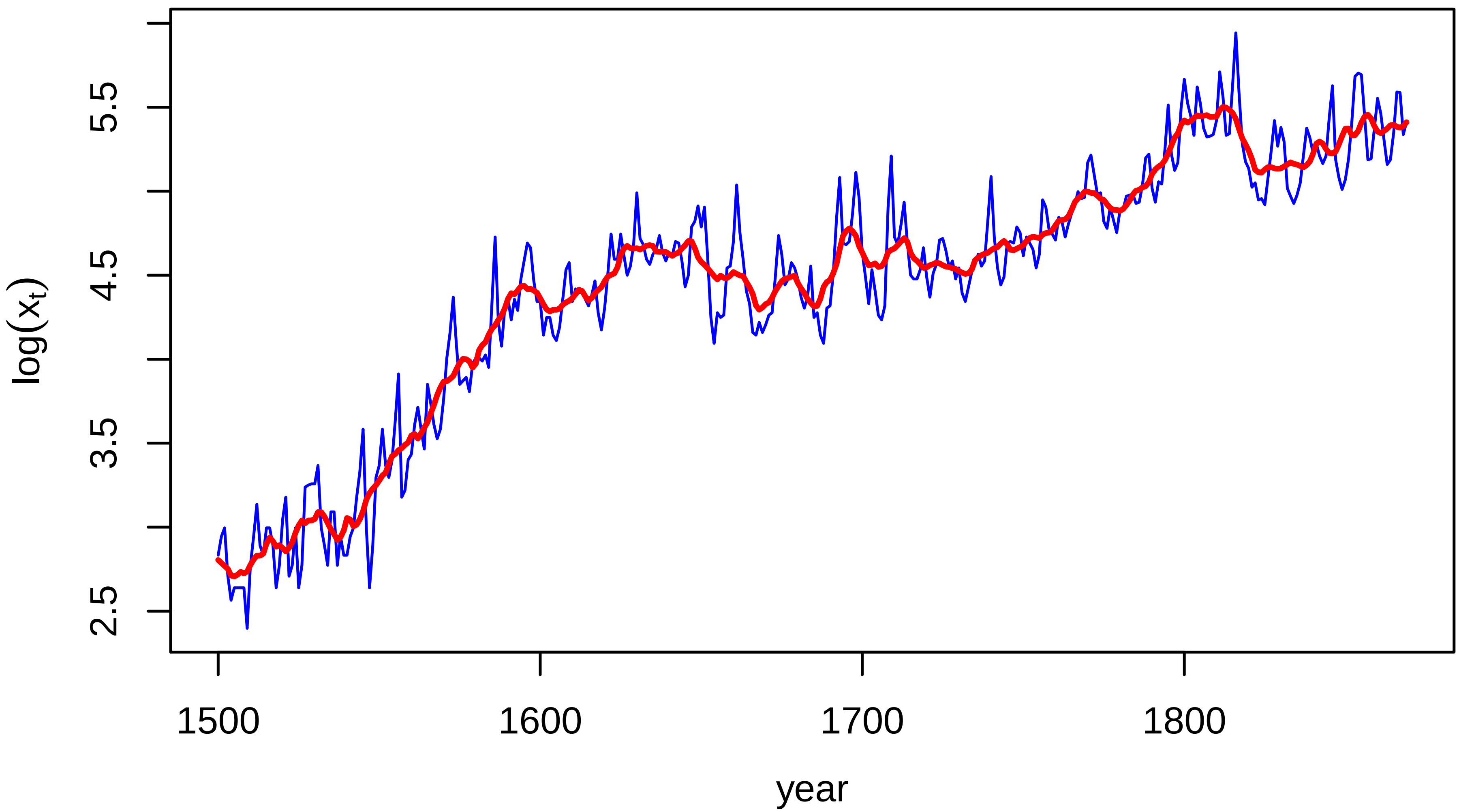
  where, by definition,

  * $x_{-q+1} = \cdots = x_0 = x_1$ and
  * $x_{n+1} = \cdots = x_{n+q} = x_n$

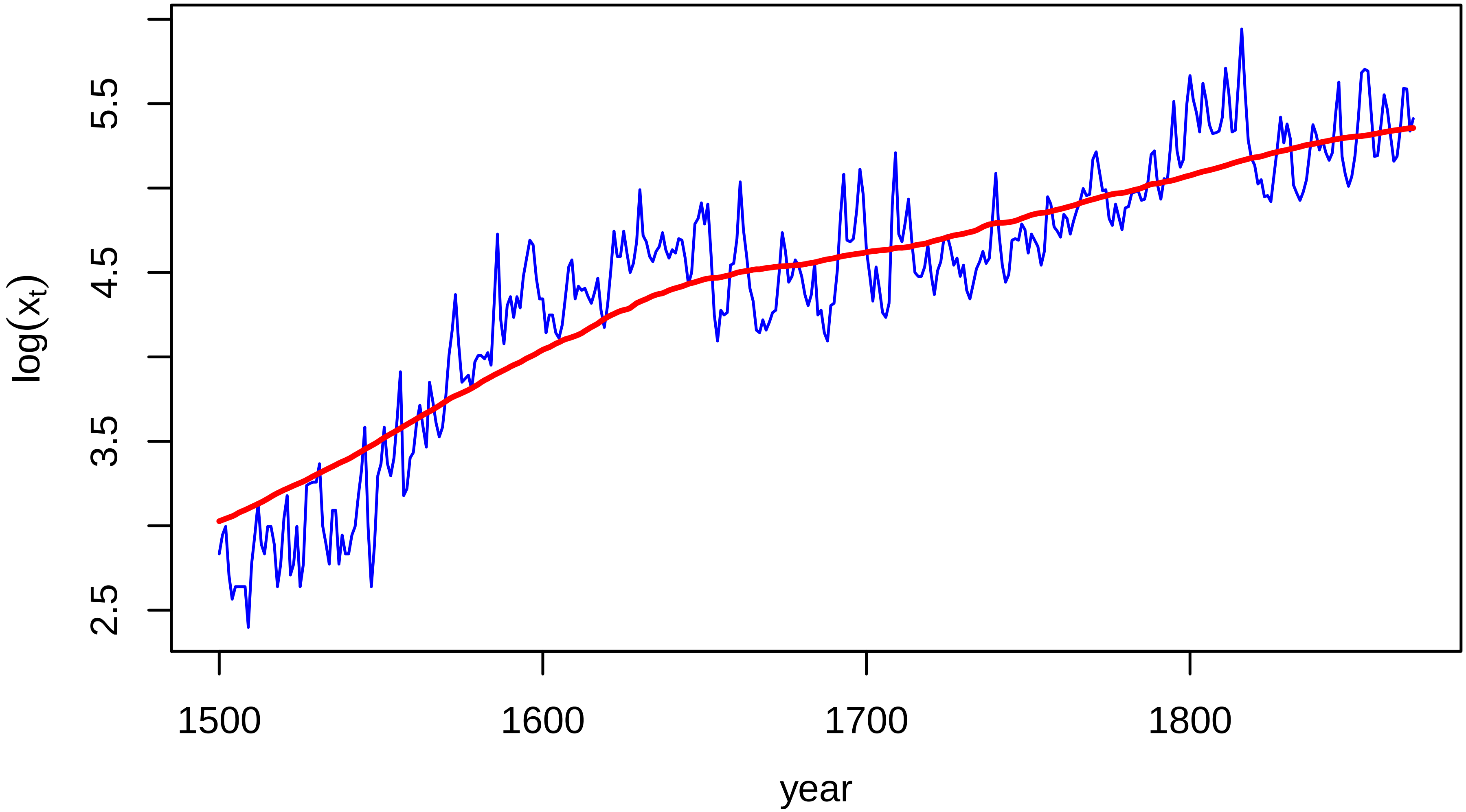
  note: possible to define these $2q$ unknown x_t's in other ways

- as examples, consider estimating $\{m_t\}$ for log of Beverage wheat price index using $q = 5$, 20 and 80

11-Term ($q = 5$) Moving Average Estimate of $\{m_t\}$



161-Term ($q = 80$) Moving Average Estimate of $\{m_t\}$



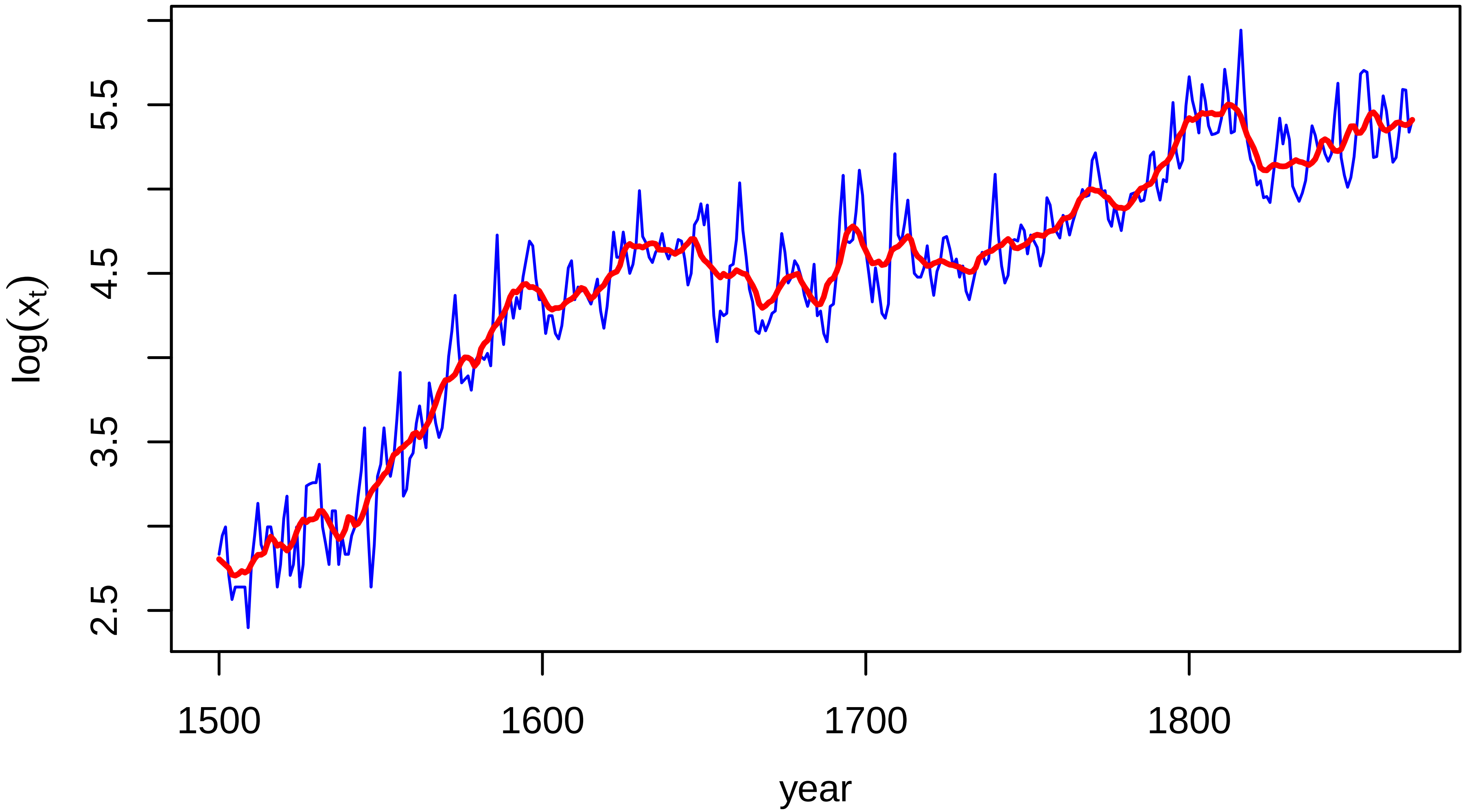
Trend Estimation via Two-Sided Filters: IV

- rather than increasing q to get more smoothing, can also apply same filter repeatedly; i.e., let

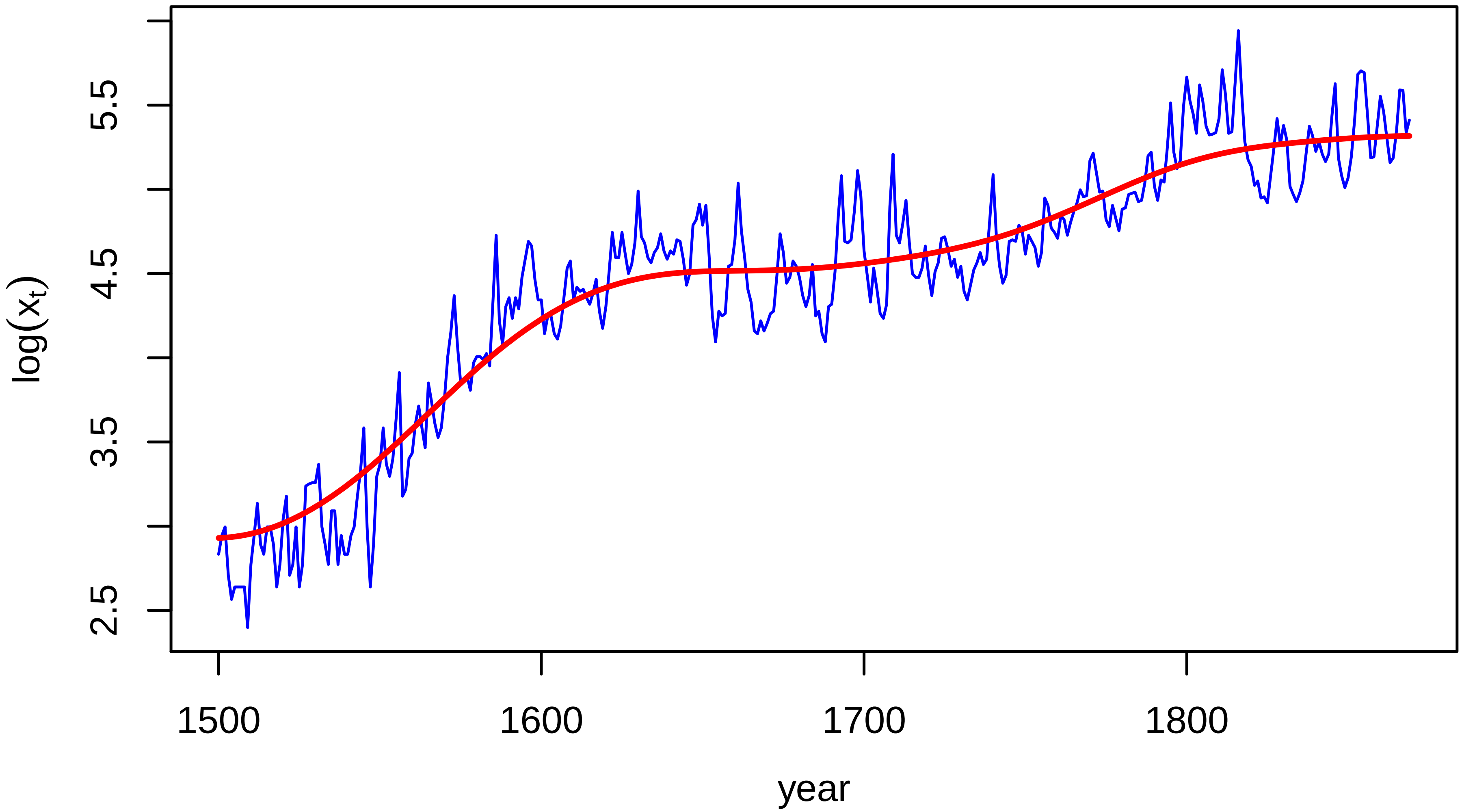
$$w_t^{(k)} = \frac{1}{2q+1} \sum_{j=-q}^{q} w_{t-j}^{(k-1)}$$

for $k = 1, 2, \ldots, K$, with $w_t^{(0)} \stackrel{\text{def}}{=} x_t$

One $(K = 1)$ Application of 11-Term MA Smoother

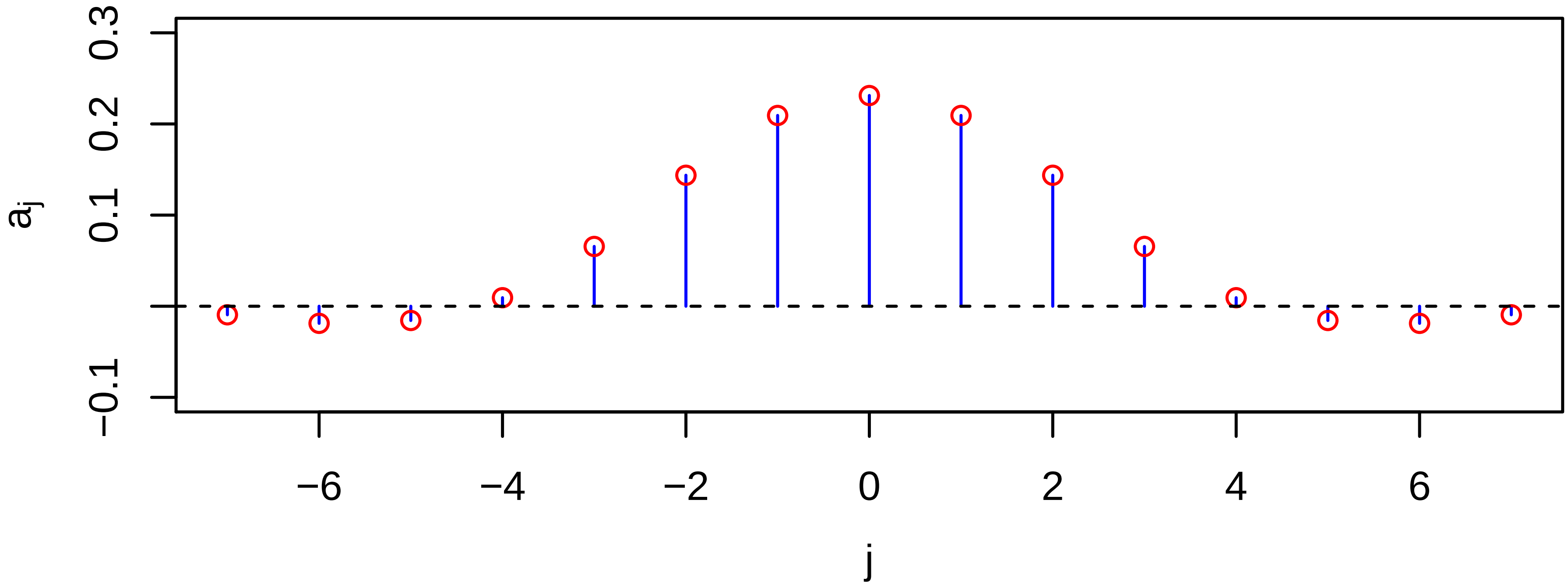


$K = 80$ Applications of 11-Term MA Smoother

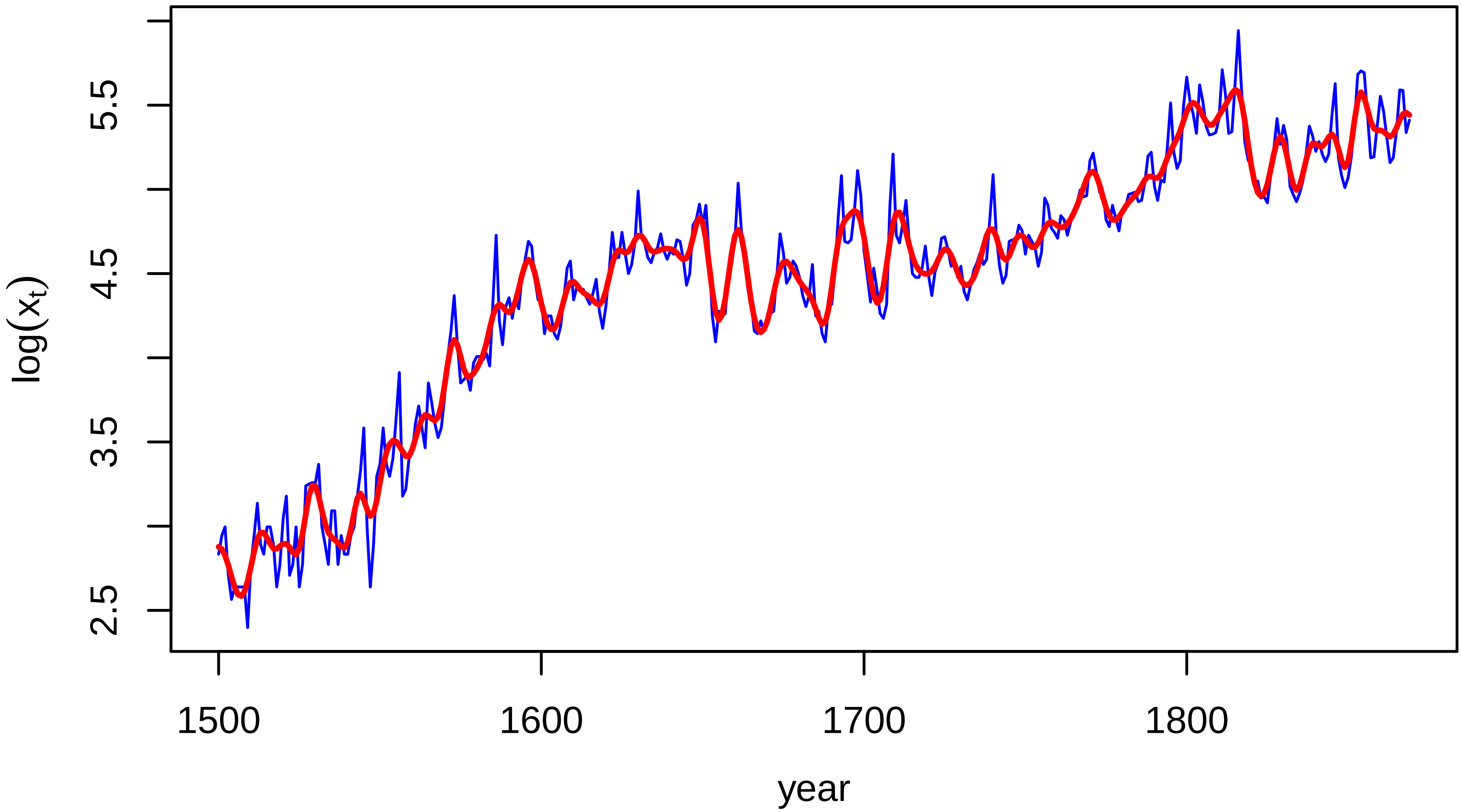


Trend Estimation via Two-Sided Filters: V

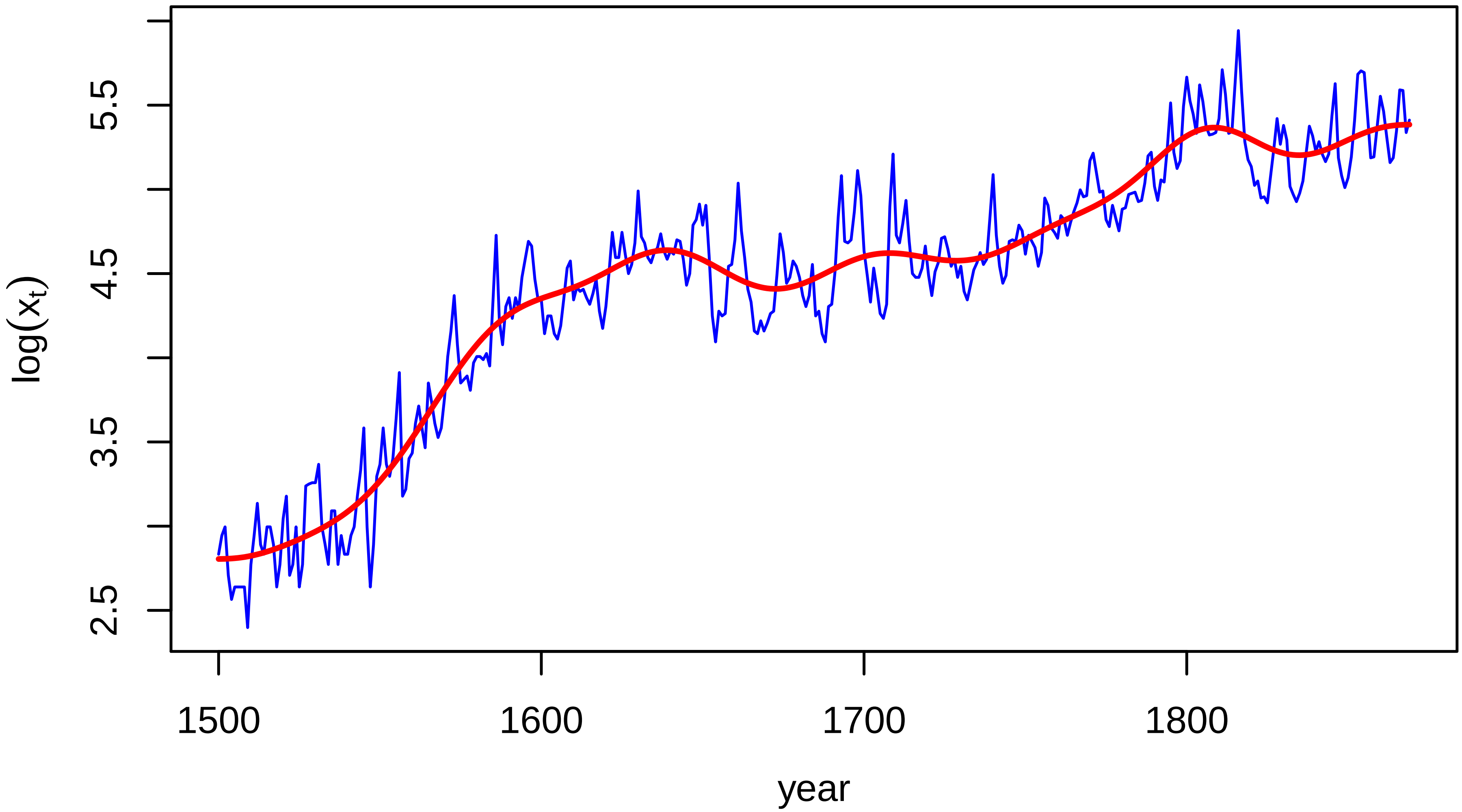
- moving average filter is an example of a smoothing filter
- *lots* of other filters can serve as smoothing filters, one example being Spencer's 15-point filter, which is designed to pass polynomials of degree 3 or less without distortion



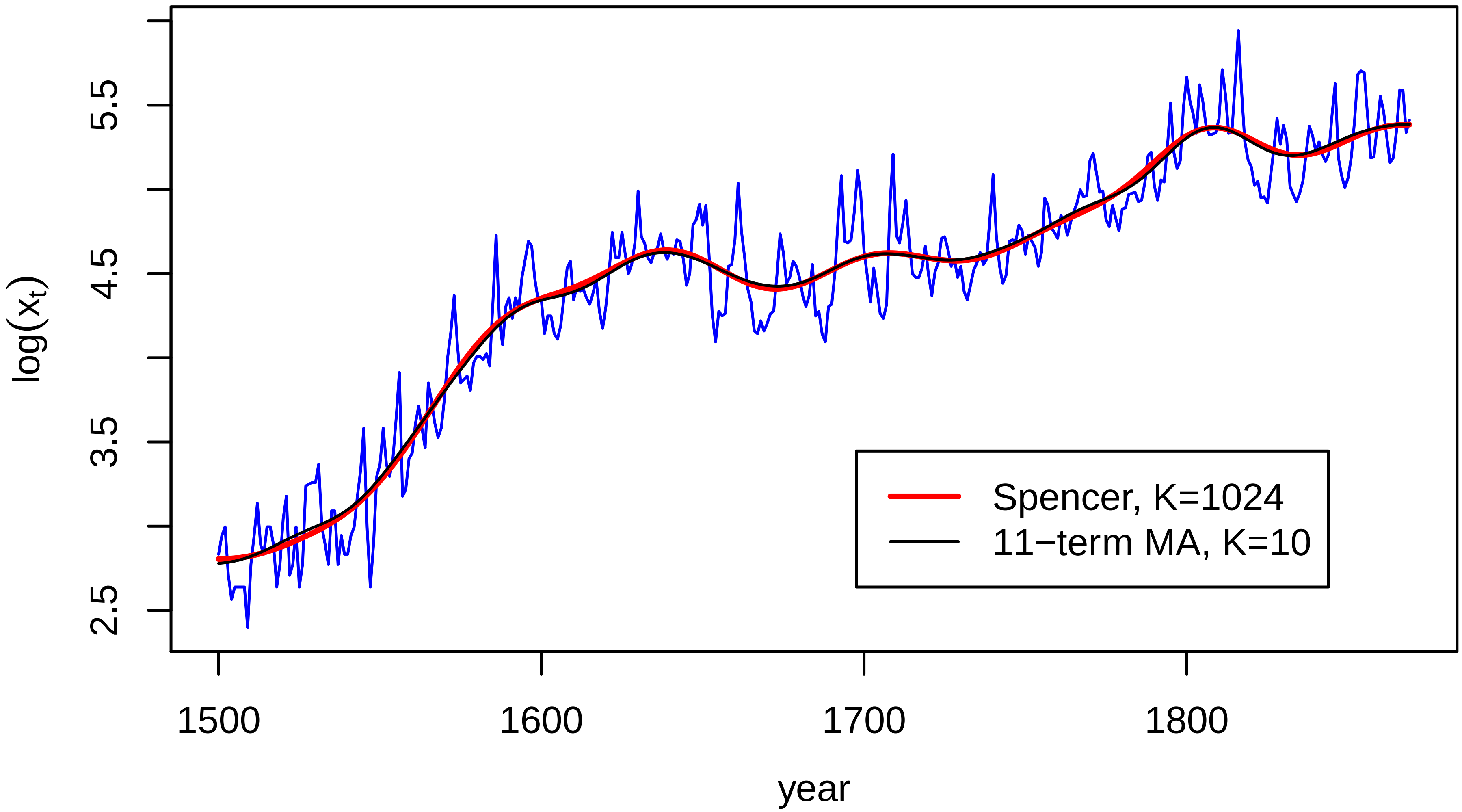
Trend Estimate Based on Spencer's 15-Point Filter



$K = 1024$ Applications of Spencer's 15-Point Filter



$K = 1024$ Applications of Spencer's 15-Point Filter



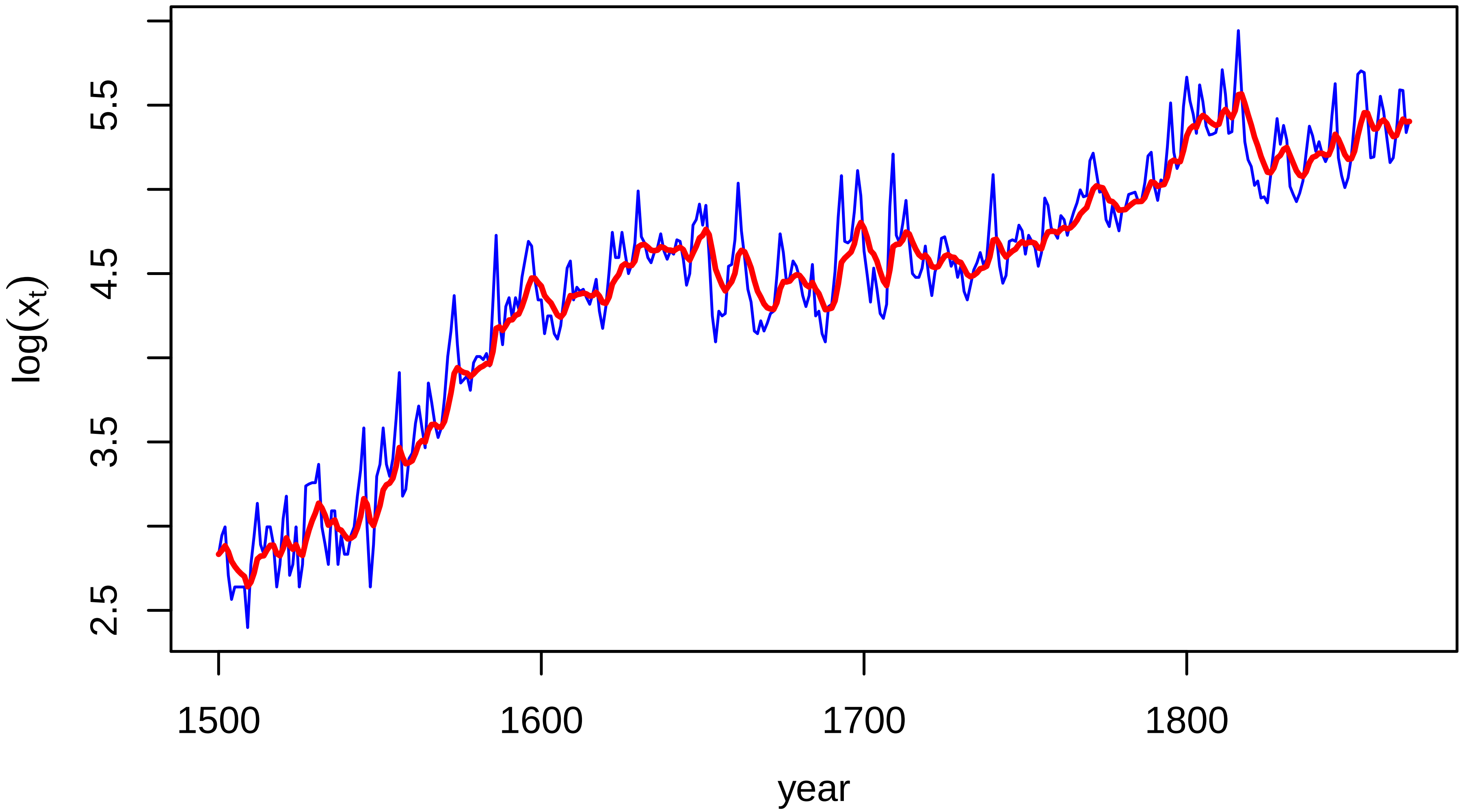
Trend Estimation via Exponential Smoothing: I

- exponential smoothing offers another way to estimate a trend
  – also called exponentially weighted moving average (EWMA)
- estimate of $\{m_t\}$ defined by the recursions

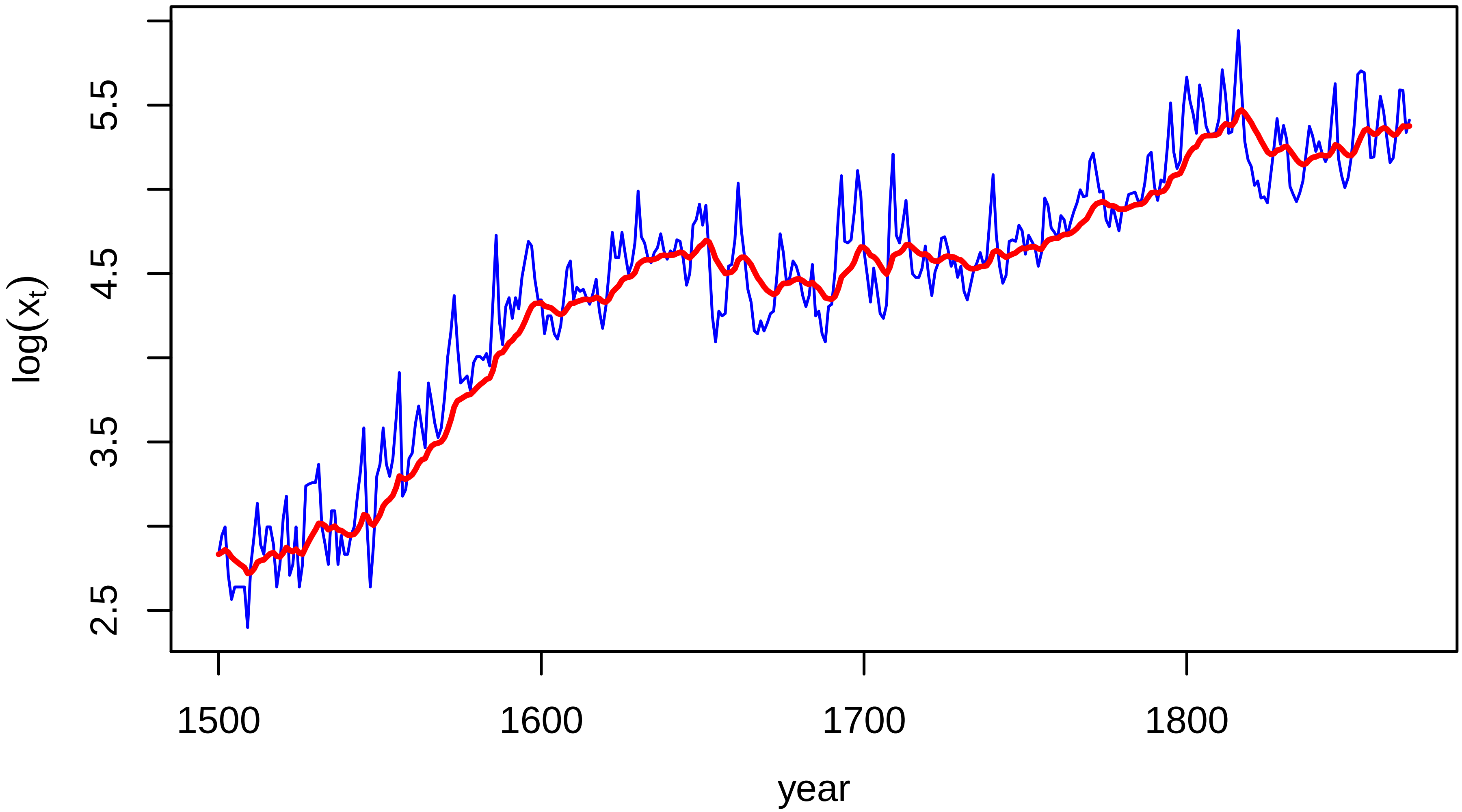
$$\hat{m}_t = \alpha x_t + (1 - \alpha)\hat{m}_{t-1}, \quad t = 2, 3, \ldots, n,$$

  with $\hat{m}_1 \stackrel{\text{def}}{=} x_1$, where $0 \leq \alpha \leq 1$
- α often chosen subjectively by trial and error (α close to 1 gives little smoothing; α close to 0 results in lots of smoothing)
- $\hat{m}_t$ at each time t only depends on $x_1, x_2, \ldots, x_t$, so this type of filter is deemed *one-sided* and *causal*
- EWMA usually introduced as a simple approach for forecasting a time series – not very appealing as trend estimator due to shifts in time, as following examples show

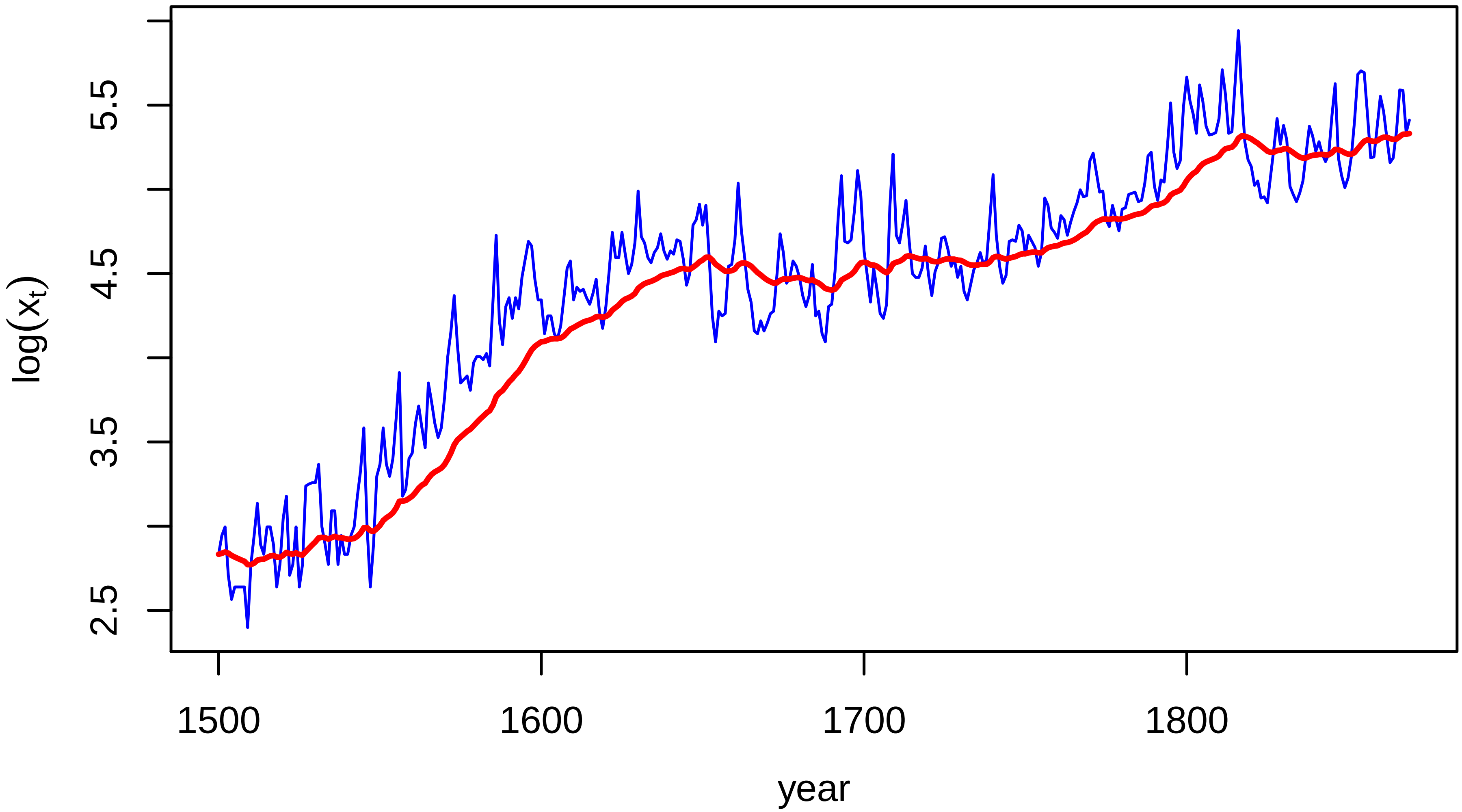
Exponential Smoothing with $\alpha = 0.2$



Exponential Smoothing with $\alpha = 0.1$



Exponential Smoothing with $\alpha = 0.05$



Trend Estimation via Exponential Smoothing: II

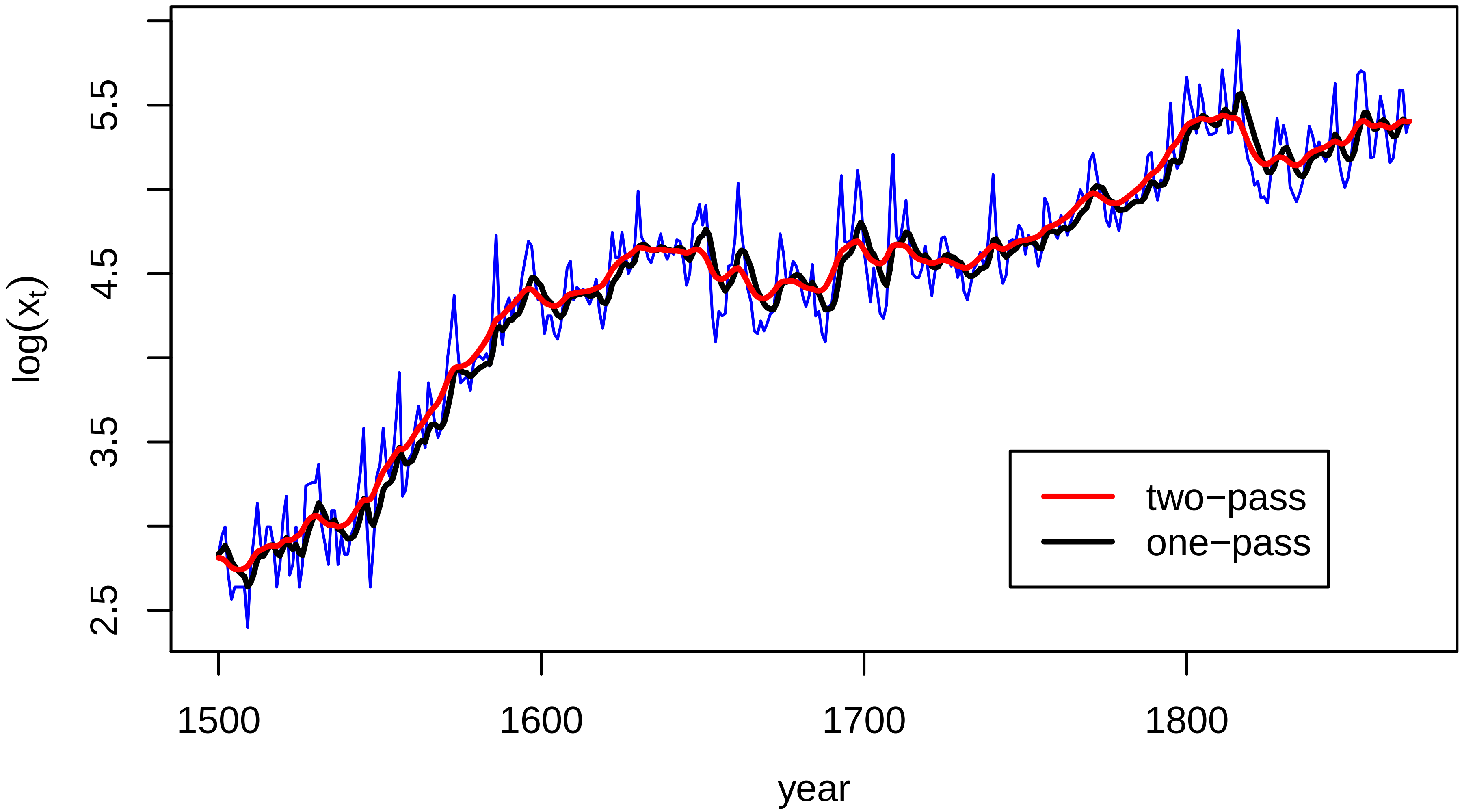
- can eliminate shifts by repeating same procedure on $\hat{m}_t$'s, but going in reverse direction; i.e.,

$$\hat{m}'_t = \alpha \hat{m}_t + (1-\alpha)\hat{m}'_{t+1}, \quad t = n-1, n-2, \ldots, 1,$$

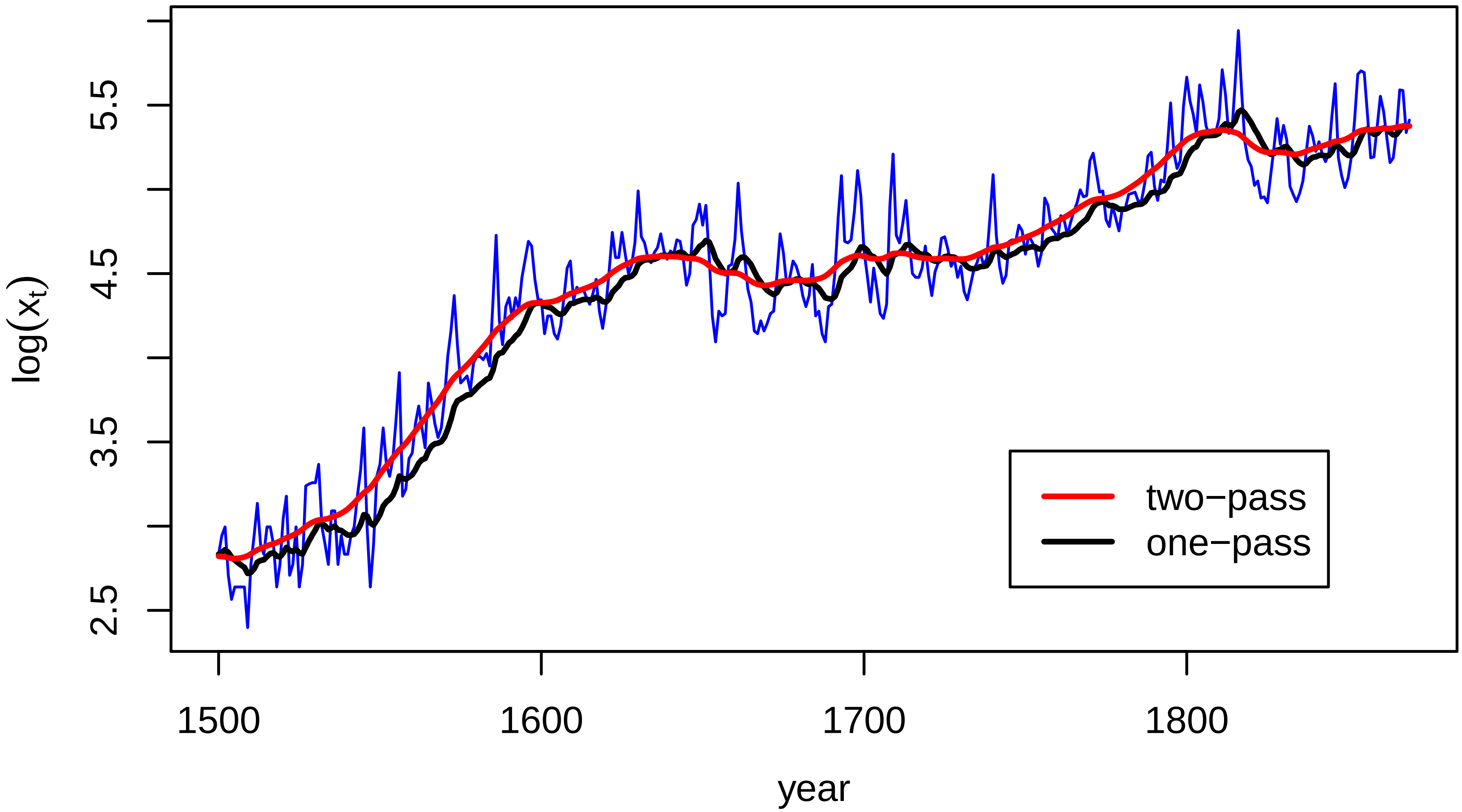
  with $\hat{m}'_n \stackrel{\text{def}}{=} \hat{m}_n$

- filtering in reverse direction is one-sided, but not causal
- let's call $\{\hat{m}'_t\}$ 'two-pass' exponential smoothing and regard original version $\{\hat{m}_t\}$ as 'one-pass'

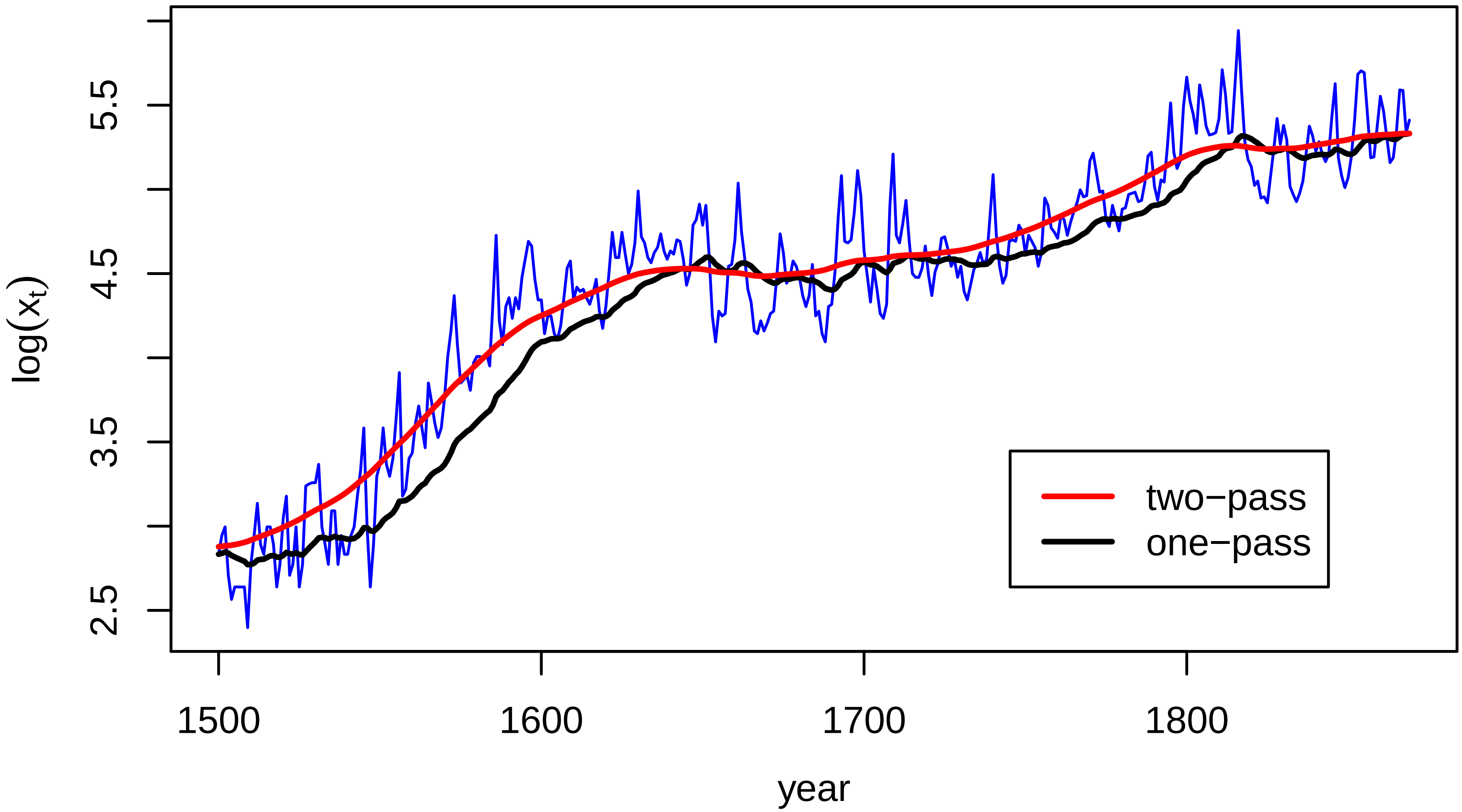
Exponential Smoothing with $\alpha = 0.2$



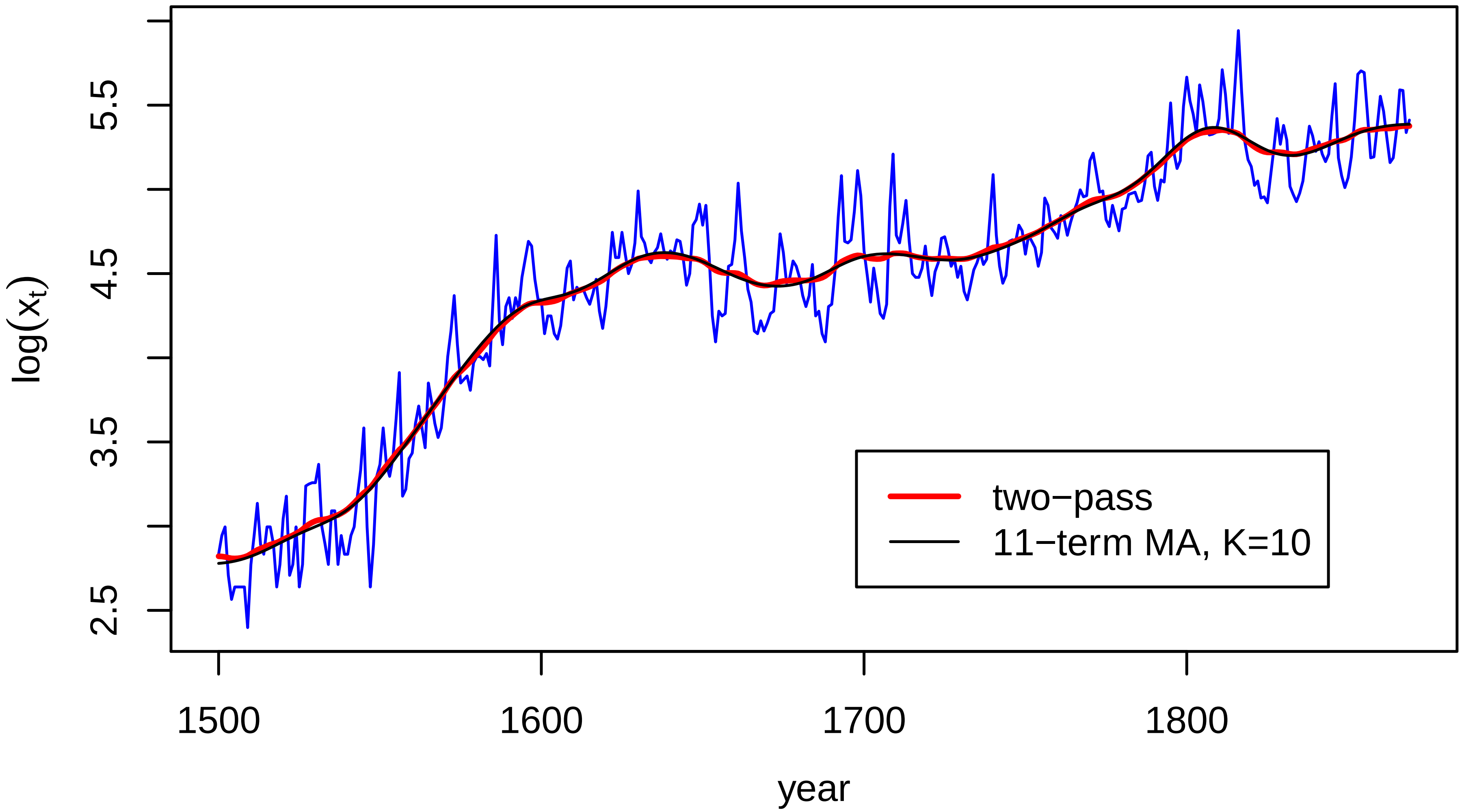
Exponential Smoothing with $\alpha = 0.1$



Exponential Smoothing with $\alpha = 0.05$



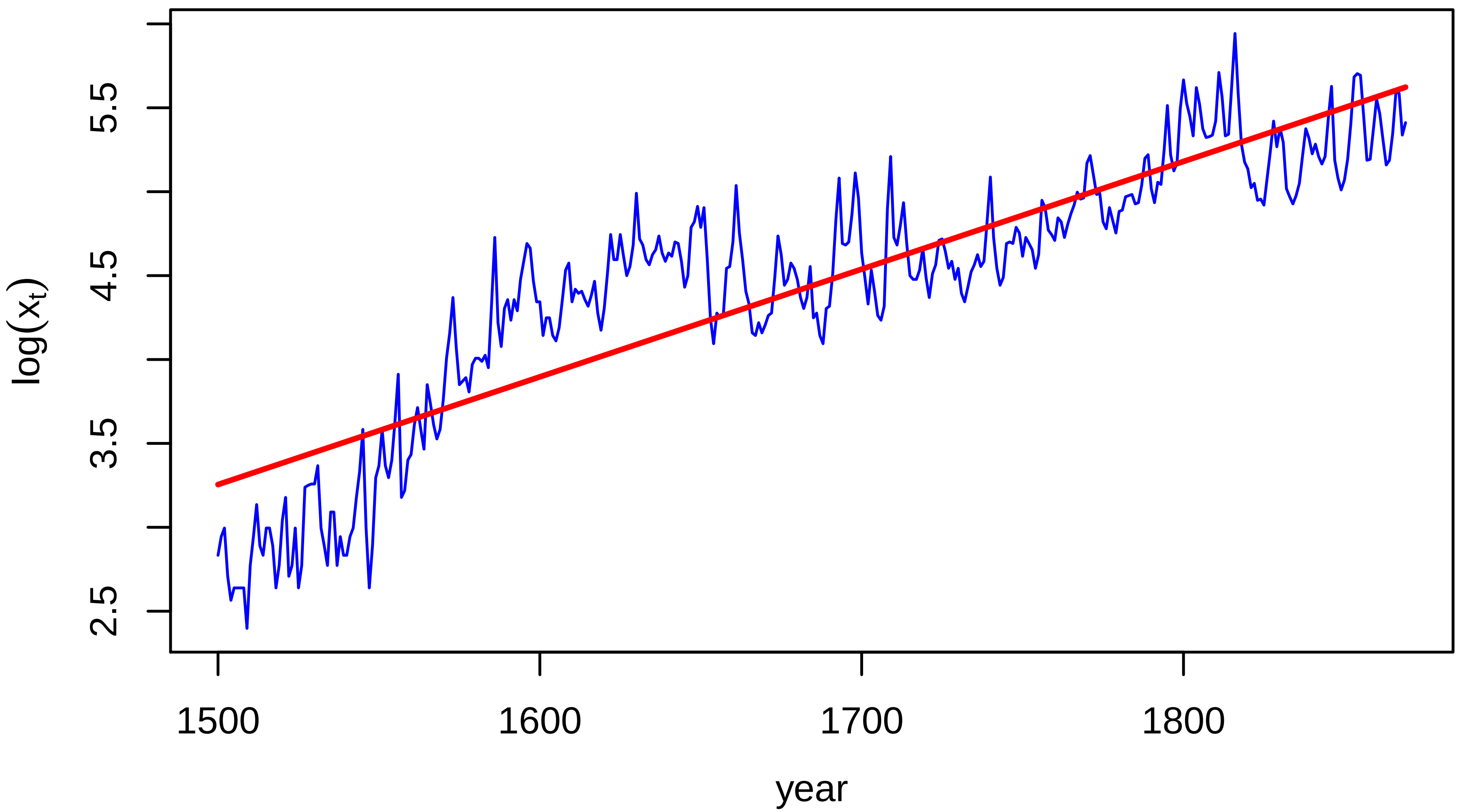
Exponential Smoothing with $\alpha = 0.1$



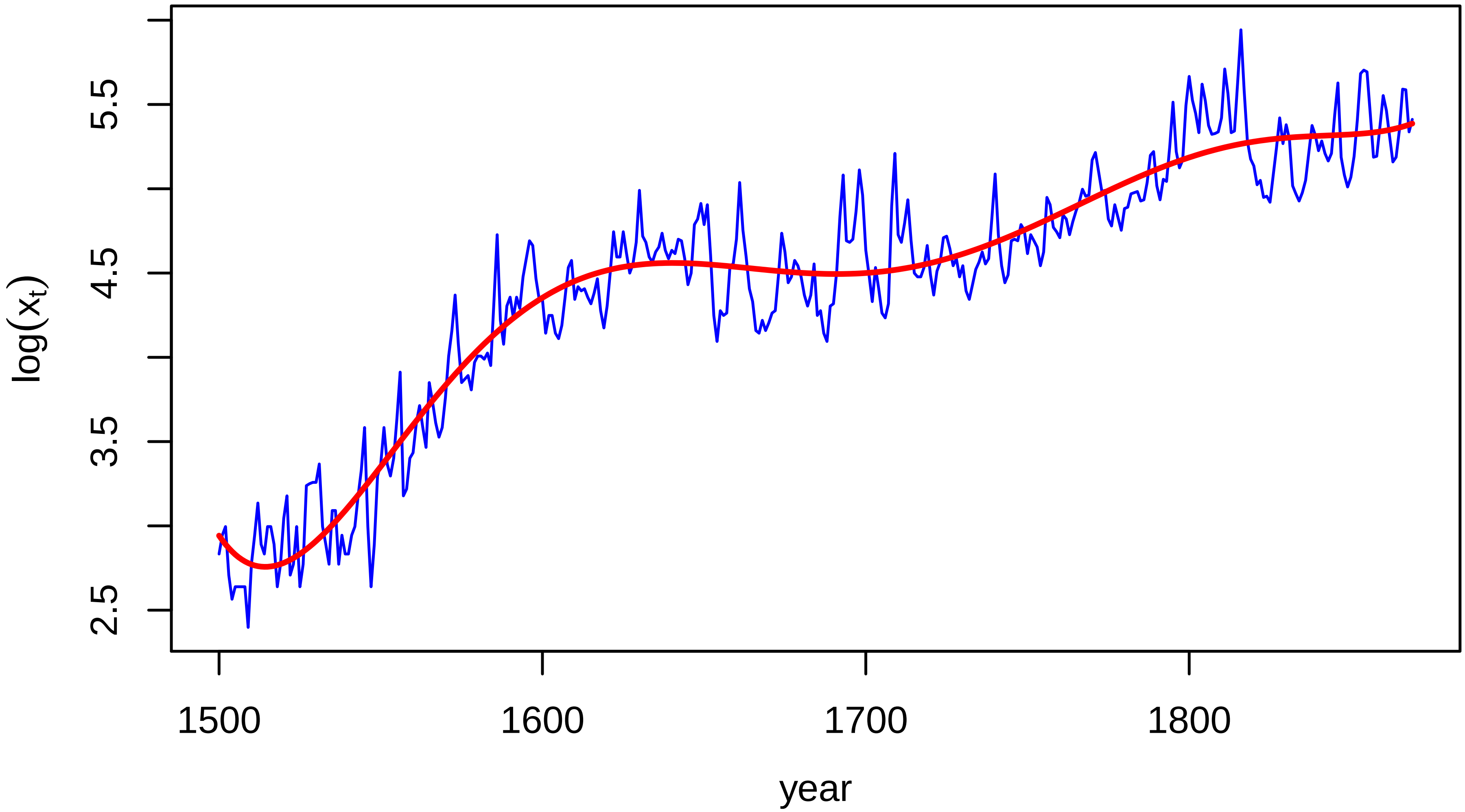
Trend Estimation via Polynomial Fitting

- have looked at linear modeling of trend Lake Huron levels (overhead III–4) and quadratic for USA population (III–7)
- can entertain polynomial trends of other orders as well: let
$$m_t = \sum_{j=0}^{k} c_j t^j,$$
where $k = 0, 1, 2, \ldots$ for constant, linear, quadratic, ...
- can estimate unknown c_j's via least squares: minimize
$$\sum_{t=1}^{n} (x_t - m_t)^2 = \sum_{t=1}^{n} \left(x_t - \sum_{j=0}^{k} c_j t^j \right)^2$$
as a function of $c_0, c_1, \ldots, c_k$
- as an example, consider log of Beveridge wheat price index

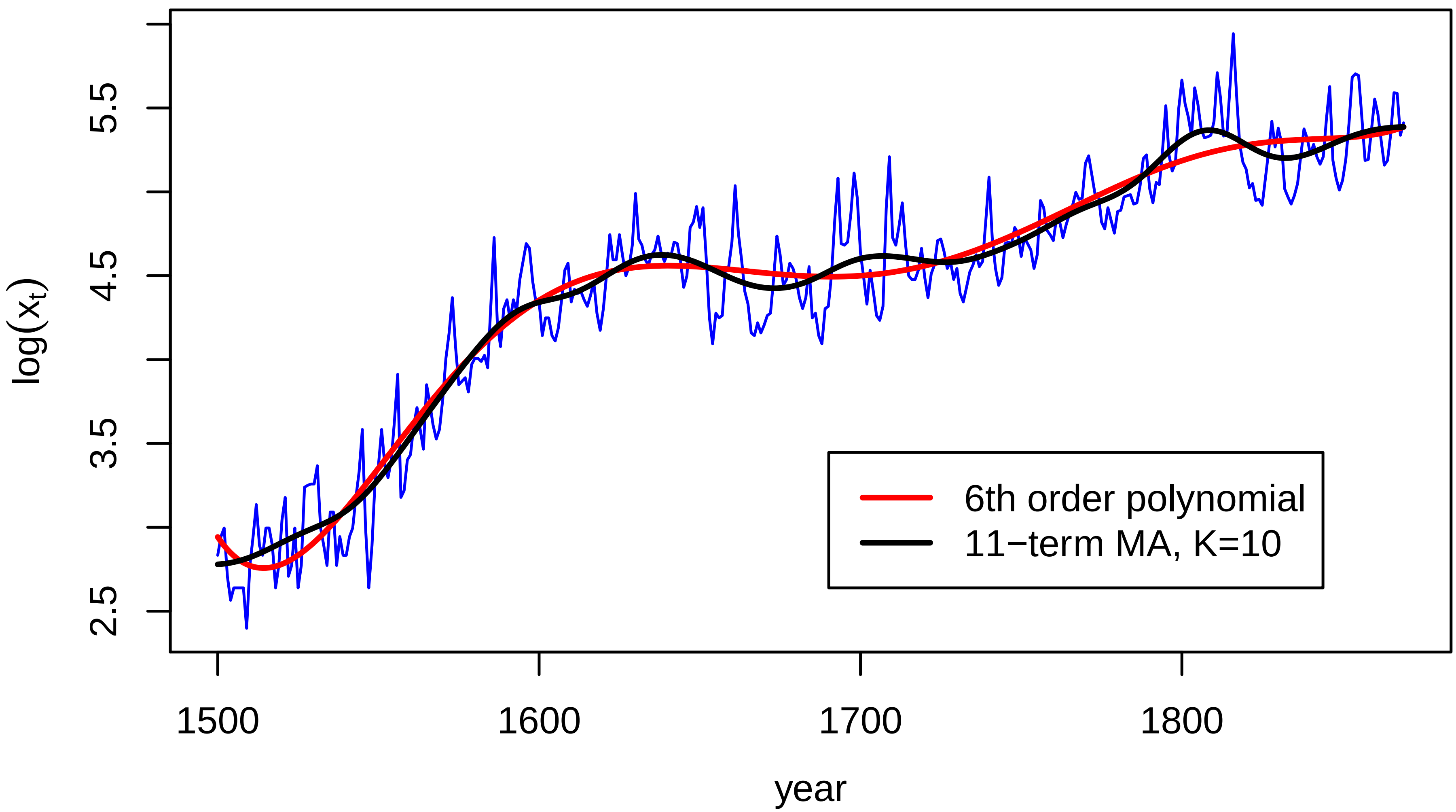
Trend Estimate Based on Fitted $c_0 + c_1 t$



Trend Estimate Based on Fitted $c_0 + c_1 t + \cdots + c_6 t^6$



Trend Estimate Based on Fitted $c_0 + c_1 t + \cdots + c_6 t^6$



Trend Estimation via Hodrick–Prescott Filter: I

- many other schemes have been proposed for trend estimation
- one such is the Hodrick–Prescott (H–P) filter, which was proposed in the economic literature in 1997 and has inspired some recent interesting research
- for a given parameter $\lambda \geq 0$, H–P estimate of trend is the sequence $\{\hat{m}_t\}$ for which, amongst all possible sequences, the two-part objective function

$$\frac{1}{2}\sum_{t=1}^{n}(x_t - \hat{m}_t)^2 + \lambda \sum_{t=2}^{n-1}(\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2$$

is minimized

  – in above, $\frac{1}{2}$ could be dropped – included in Kim et al. (2009) evidently to simply other equations

Trend Estimation via Hodrick–Prescott Filter: II

- first part
$$\frac{1}{2}\sum_{t=1}^{n}(x_t - \hat{m}_t)^2$$
quantifies fidelity: we want the trend estimate to faithfully track our time series; i.e., we want the residuals $x_t - \hat{m}_t$ to be small
- the above is small when $\{\hat{m}_t\}$ is faithful to $\{x_t\}$
- note that, if we set $\lambda = 0$ so that the objective function is just the above, then $\{\hat{m}_t\}$ must be the same as $\{x_t\}$ (the sum of squares is zero, the smallest possible value) – highest degree of faithfulness possible!

Trend Estimation via Hodrick–Prescott Filter: III

- second part
$$\lambda \sum_{t=2}^{n-1} (\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2$$
quantifies how smooth $\{\hat{m}_t\}$ is: trend is usually thought of as slowly varying, and hence we want it to be smooth
- the above is small when $\{\hat{m}_t\}$ is smooth
- to see why, suppose $\hat{m}_t = a + bt$, i.e., trend is linear (quite smooth! – its 2nd derivative is 0), in which case
$$\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1} = a + b(t+1) - 2a - 2bt + a + b(t-1) = 0$$
and hence
$$\lambda \sum_{t=2}^{n-1} (\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2 = 0$$

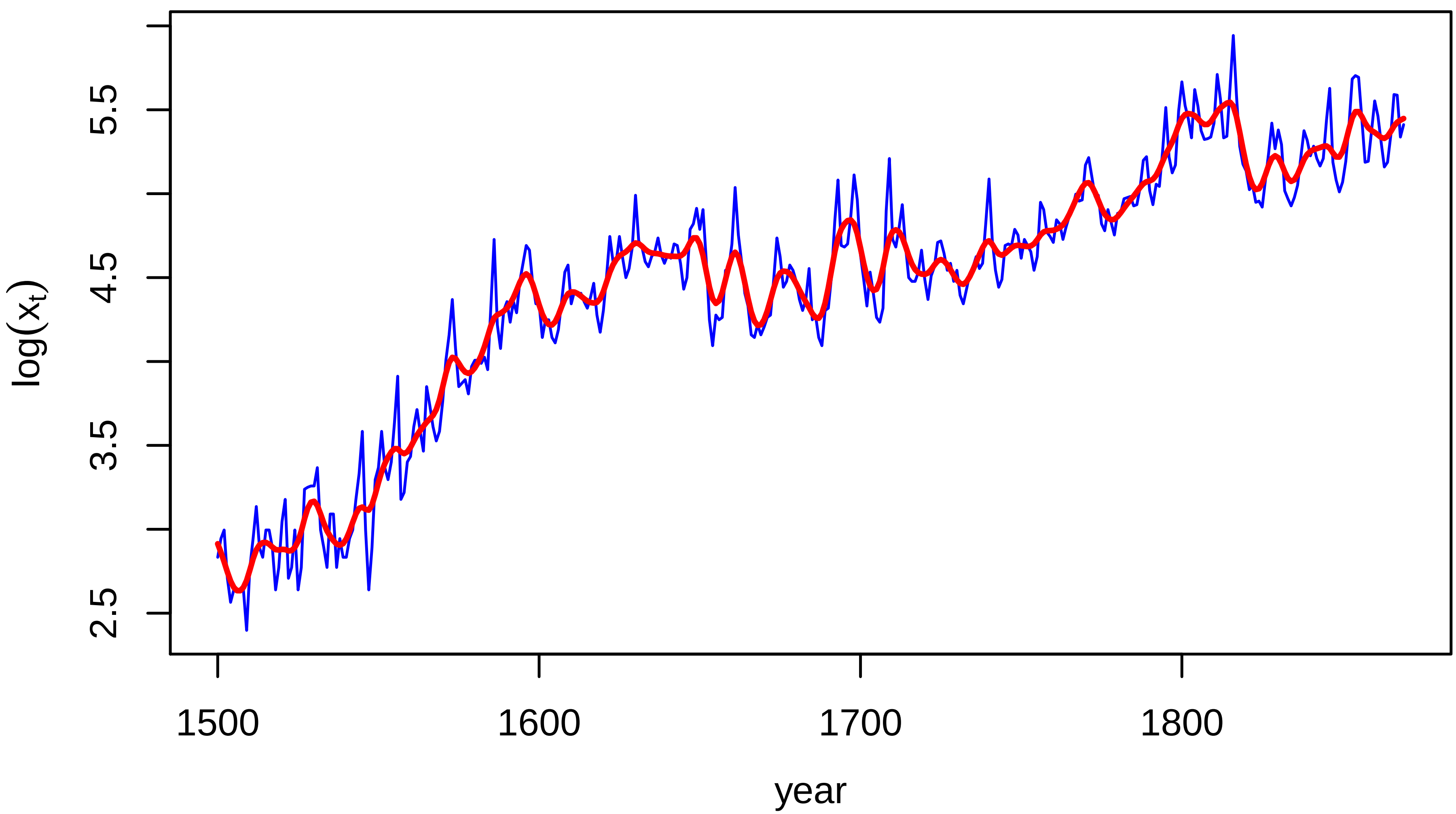
Trend Estimation via Hodrick–Prescott Filter: IV

- in general, fidelity and smoothness are in conflict
  - insisting the trend be smooth (e.g., just a line) can result in $\{\hat{m}_t\}$ not being faithful to $\{x_t\}$
  - insisting the trend be faithful (nearly the same as $\{x_t\}$) can result in $\{\hat{m}_t\}$ not being smooth
- choosing $\{\hat{m}_t\}$ such that

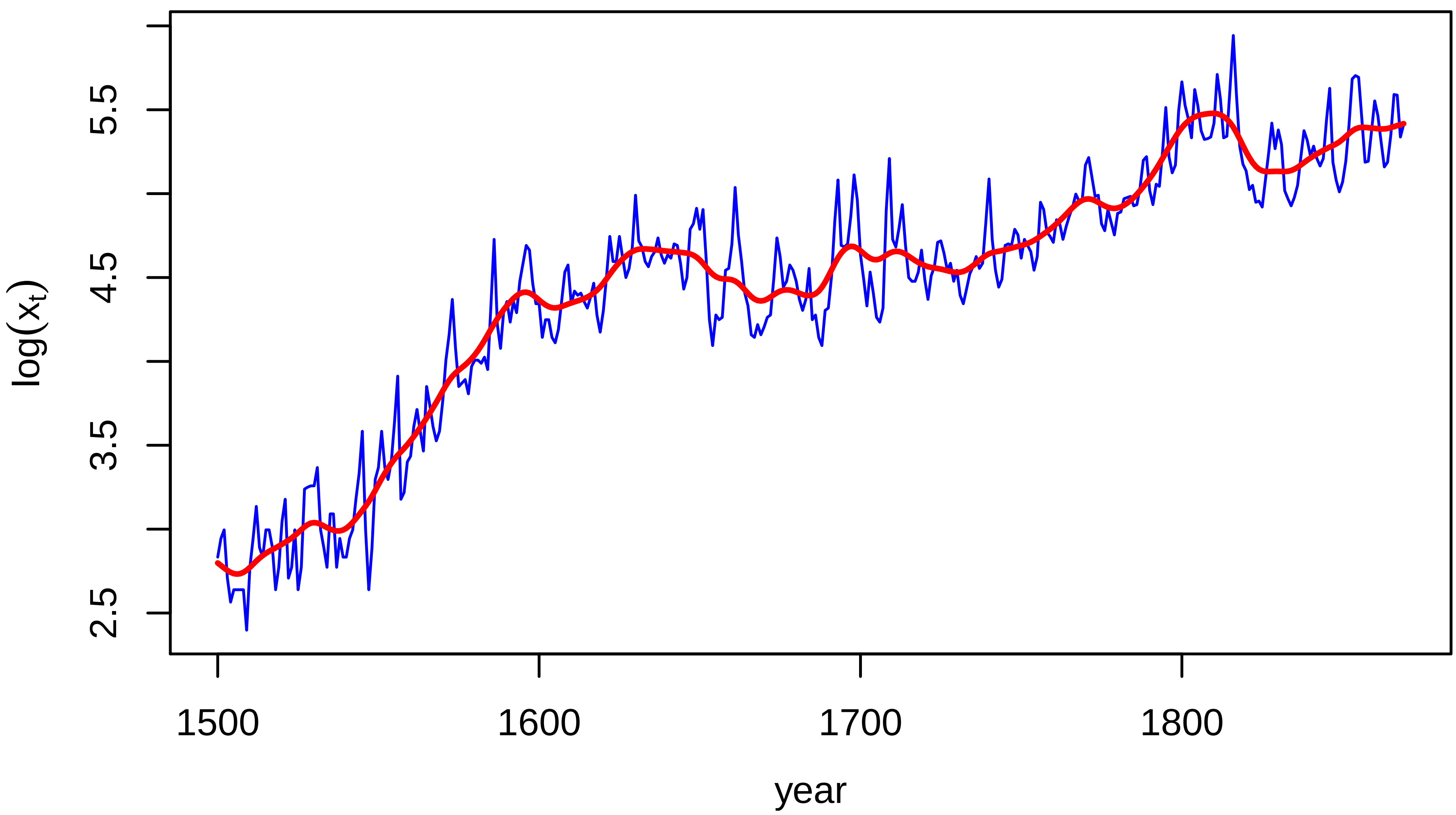
$$\frac{1}{2}\sum_{t=1}^{n}(x_t - \hat{m}_t)^2 + \lambda \sum_{t=2}^{n-1}(\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2$$

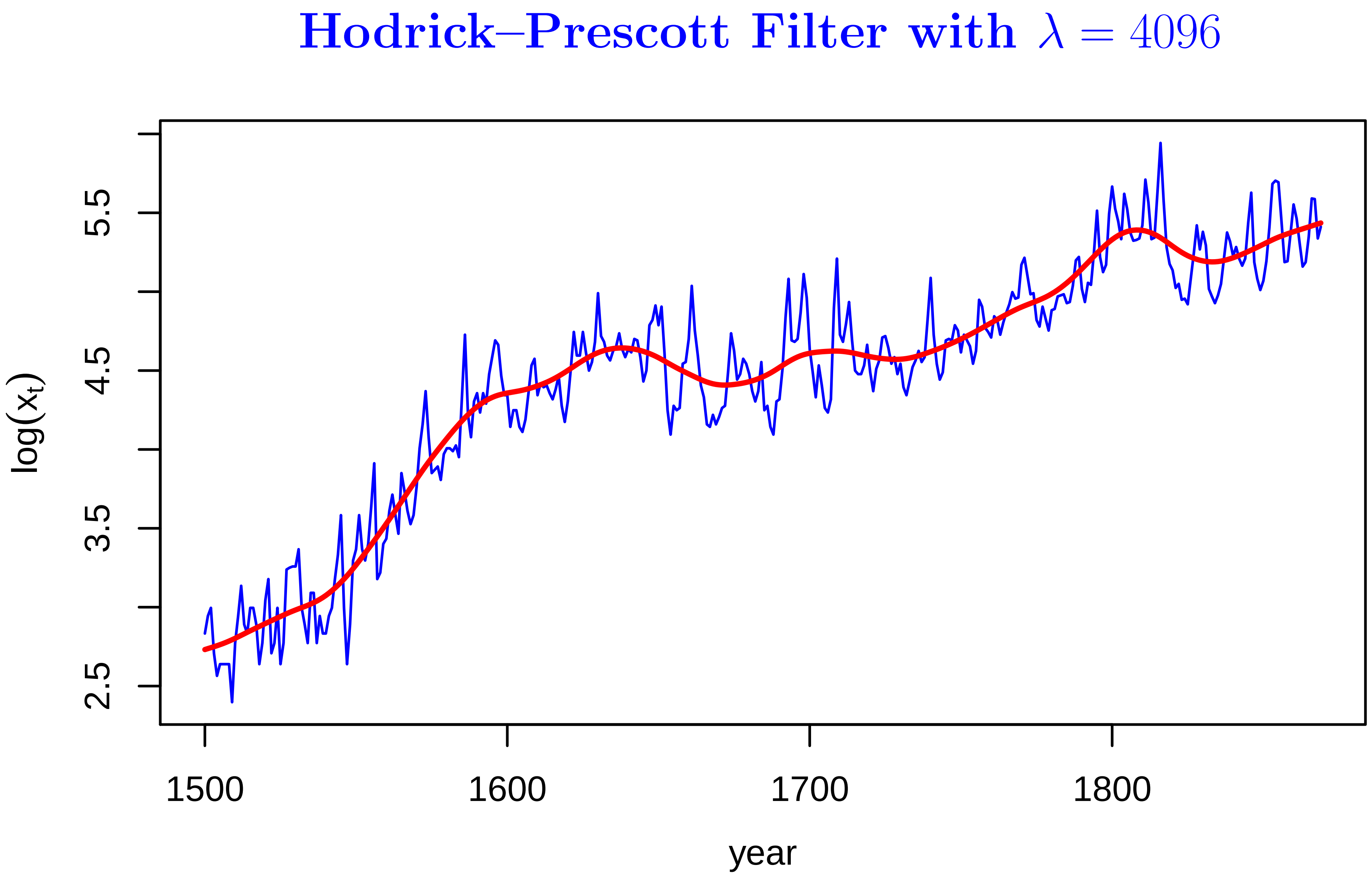
  is minimized is an attempt to strike a balance between fidelity and smoothness, with λ controlling the balance

Hodrick–Prescott Filter with $\lambda = 16$

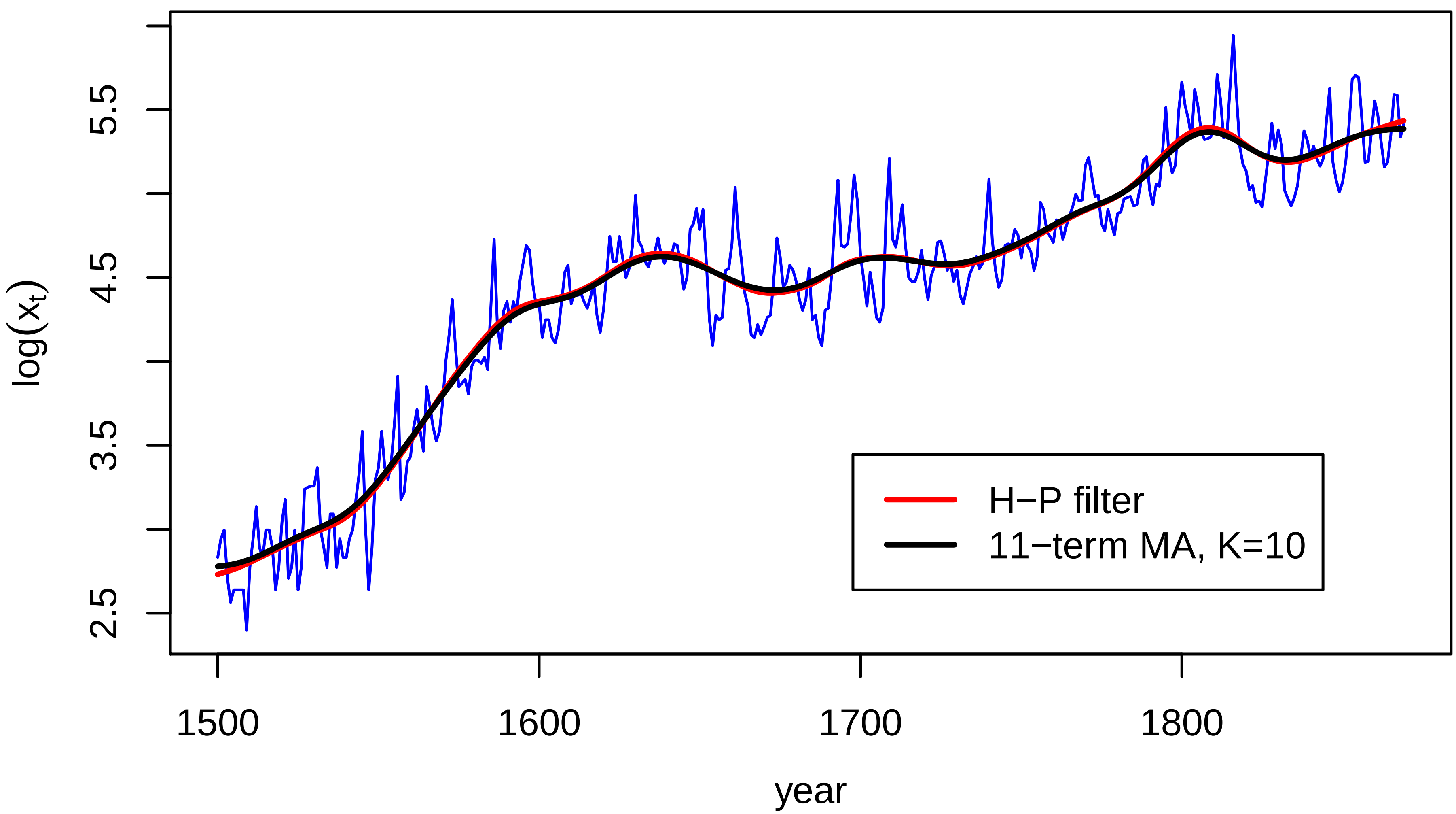


Hodrick–Prescott Filter with $\lambda = 256$




Hodrick–Prescott Filter with $\lambda = 4096$
log(x_t)
2.5
3.5
4.5
5.5
1500
1600
1700
1800
year

Hodrick–Prescott Filter with $\lambda = 4096$



Trend Estimation via Hodrick–Prescott Filter: V

- H–P filter is a linear filter because in fact

$$\hat{\boldsymbol{m}} = \left(I + 2\lambda D^T D\right)^{-1} \boldsymbol{x},$$

where $\hat{\boldsymbol{m}}$ is a column vector containing $\hat{m}_1, \ldots, \hat{m}_n$; I is the $n \times n$ identity matrix; D is an $(n-2) \times n$ matrix, namely,

$$D = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

(entries not shown above are zeros); D^T is the transpose of D; A^{-1} is the inverse of $A = I + 2\lambda D^T D$; and $\boldsymbol{x}$ is a column vector with $x_1, \ldots, x_n$

ℓ_1 Trend Filtering: I

- H–P filter inspired interesting alternate called ℓ_1 trend filtering (Kim et al., 2009; Tibshirani, 2014)
- rather than choosing $\{\hat{m}_t\}$ such that

$$\frac{1}{2}\sum_{t=1}^{n}(x_t - \hat{m}_t)^2 + \lambda \sum_{t=2}^{n-1}(\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1})^2$$

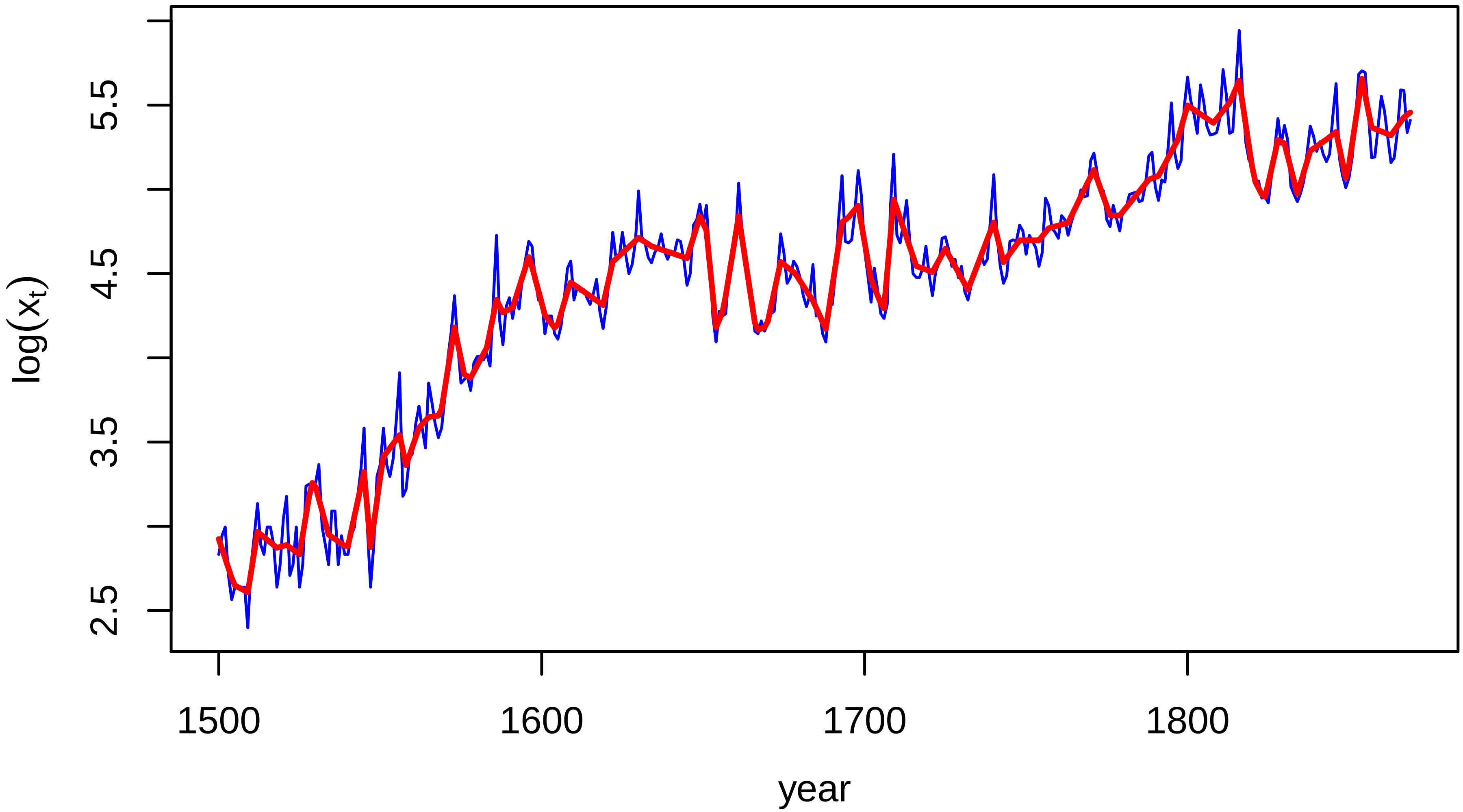
is minimized, ℓ_1 trend filtering chooses $\{\hat{m}_t\}$ such that

$$\frac{1}{2}\sum_{t=1}^{n}(x_t - \hat{m}_t)^2 + \lambda \sum_{t=2}^{n-1}|\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}|$$

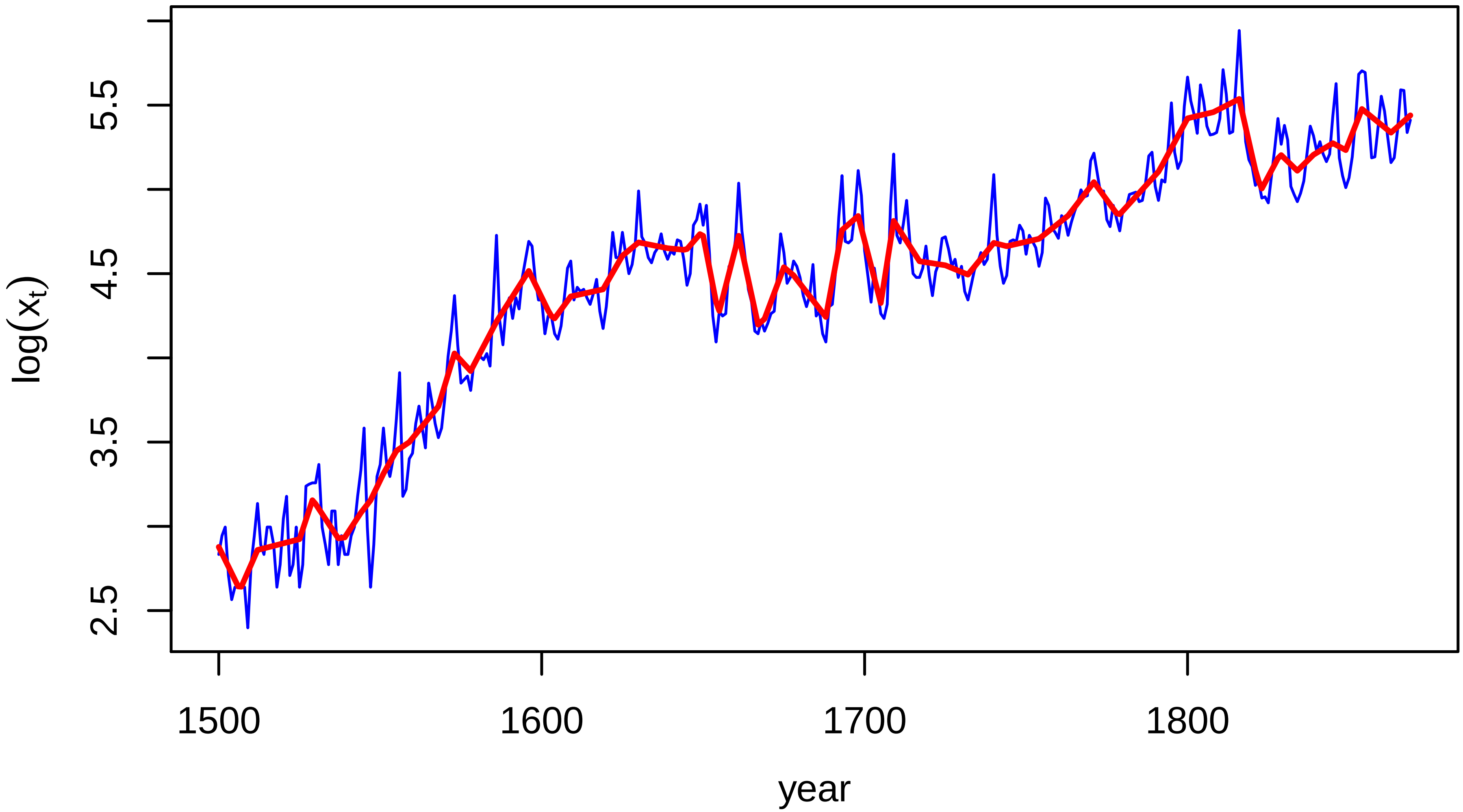
is minimized (note: setting $\lambda = 0$ again yields $\hat{m}_t = x_t$)

- resulting $\{\hat{m}_t\}$ is piecewise linear, but in general *cannot* be written as a linear transform of $\{x_t\}$

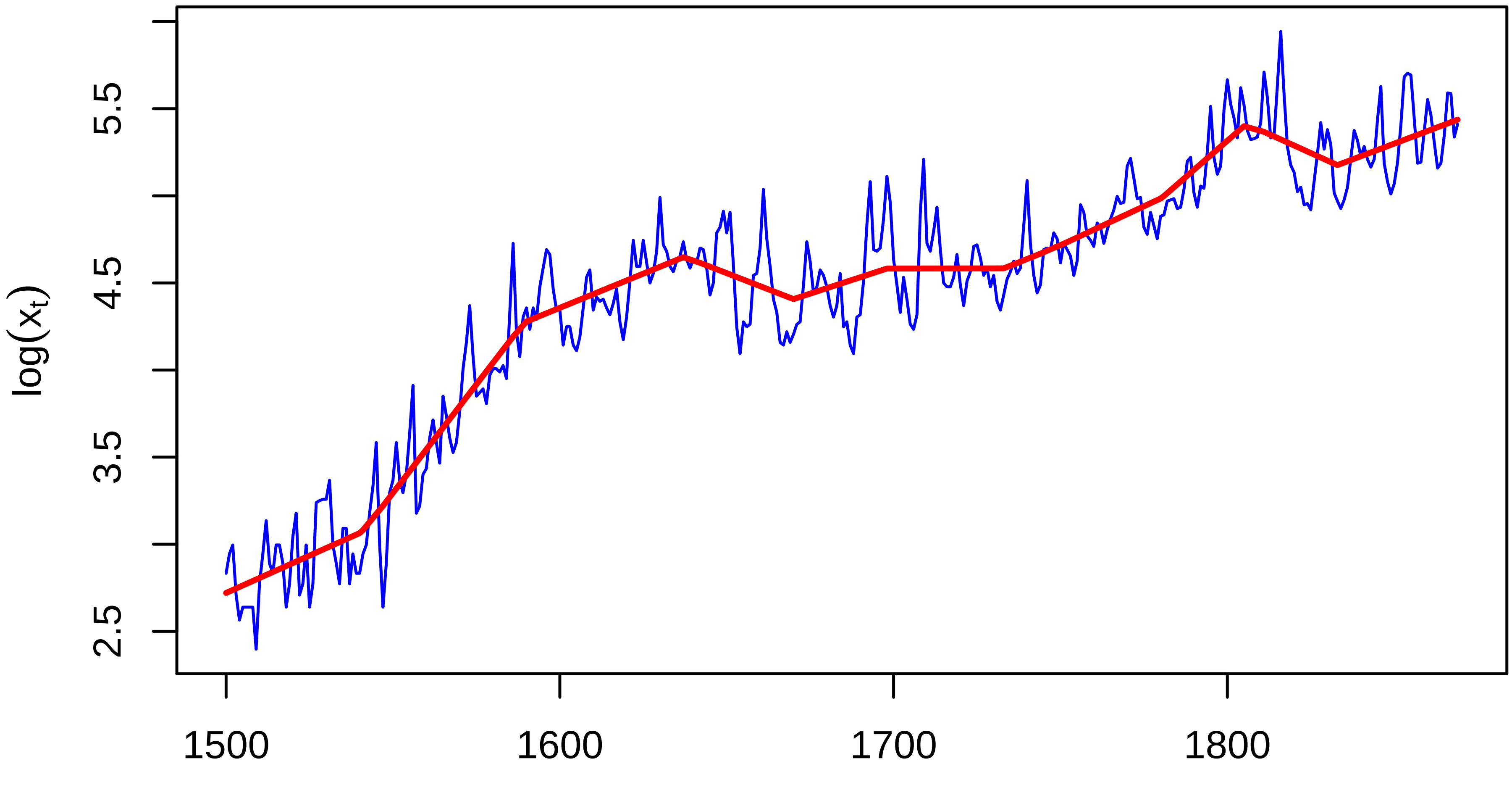
ℓ_1 Trend Filter with $\lambda = 0.2$



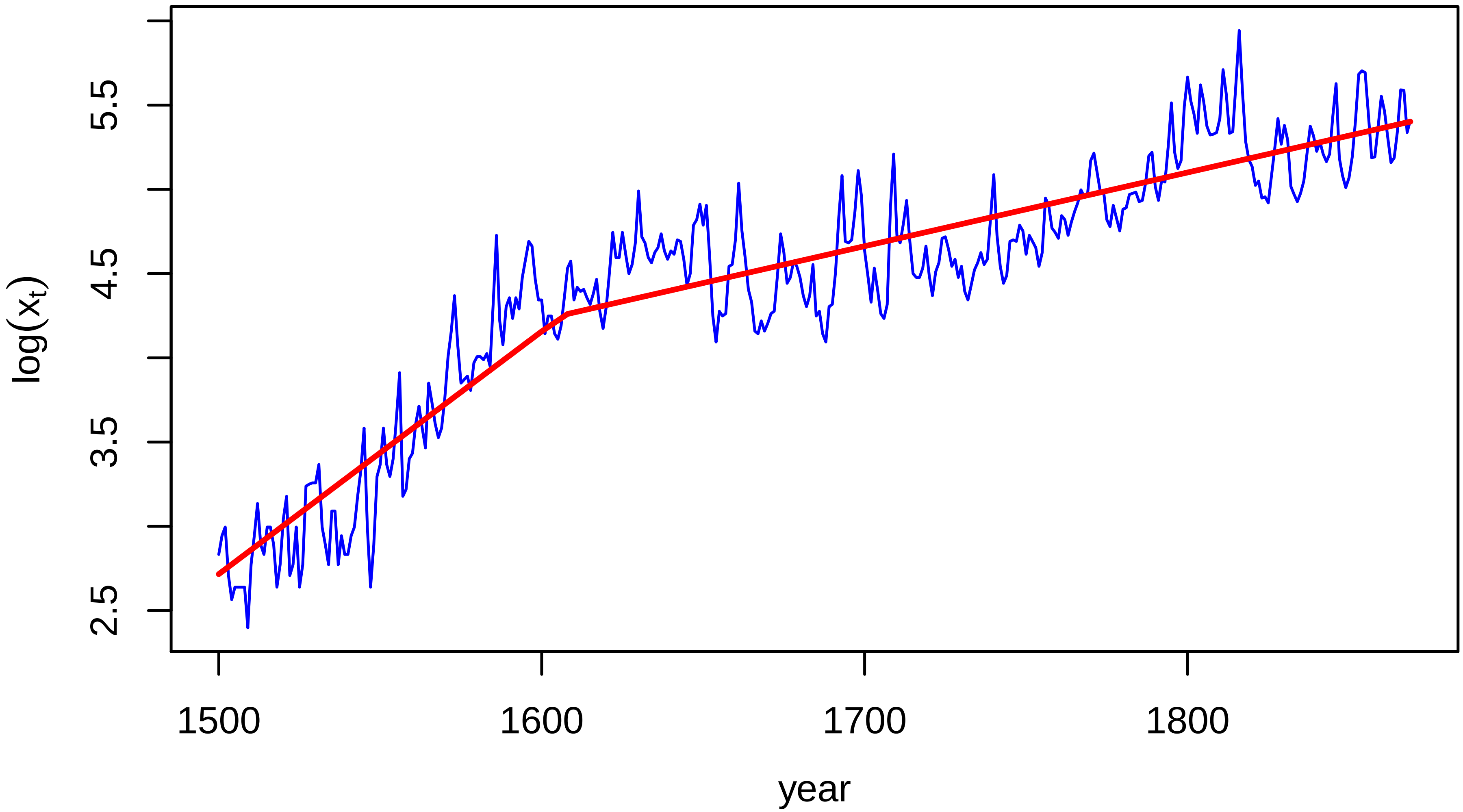
ℓ_1 Trend Filter with $\lambda = 0.5$

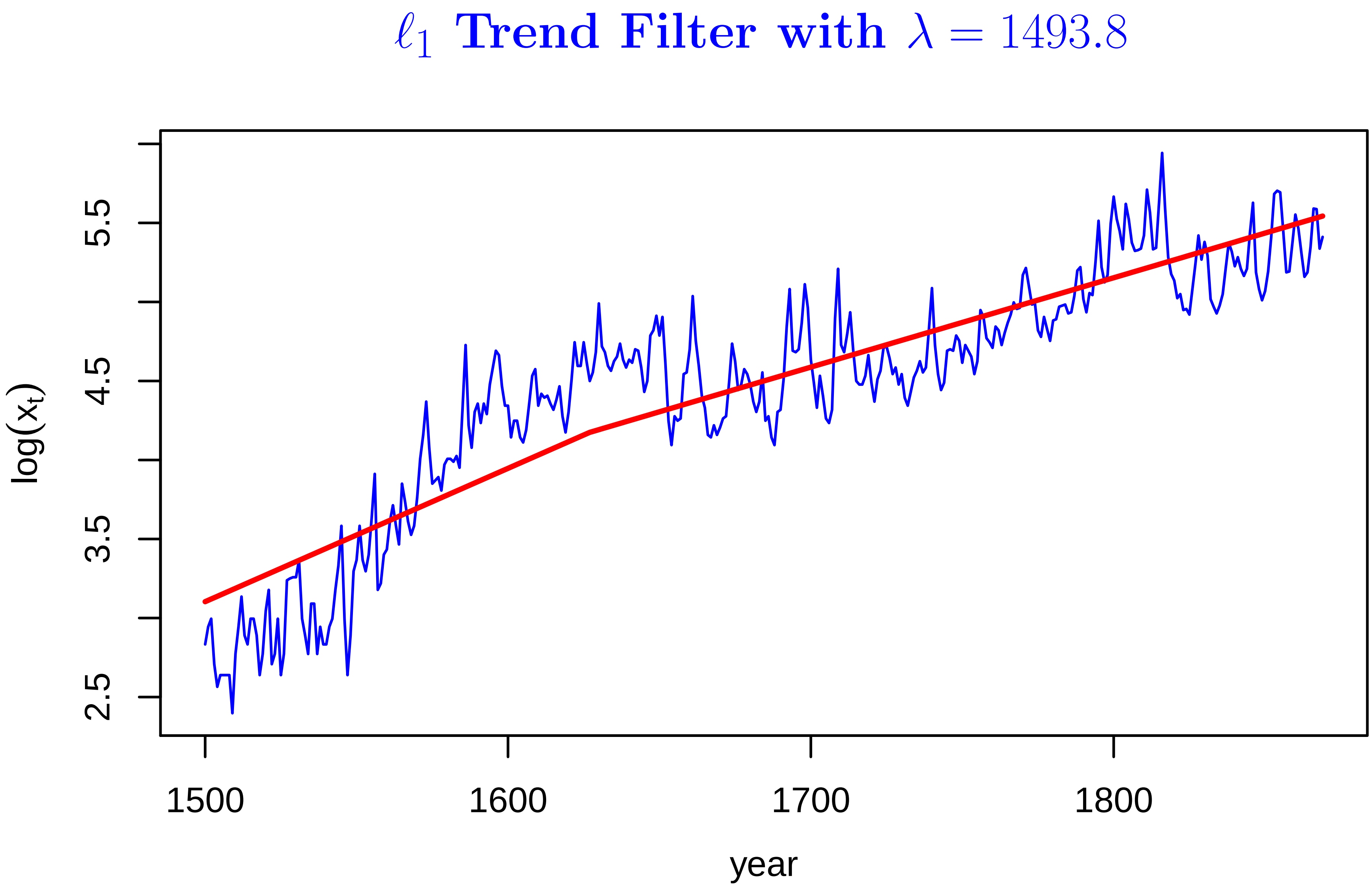


ℓ_1 Trend Filter with $\lambda = 6.8$



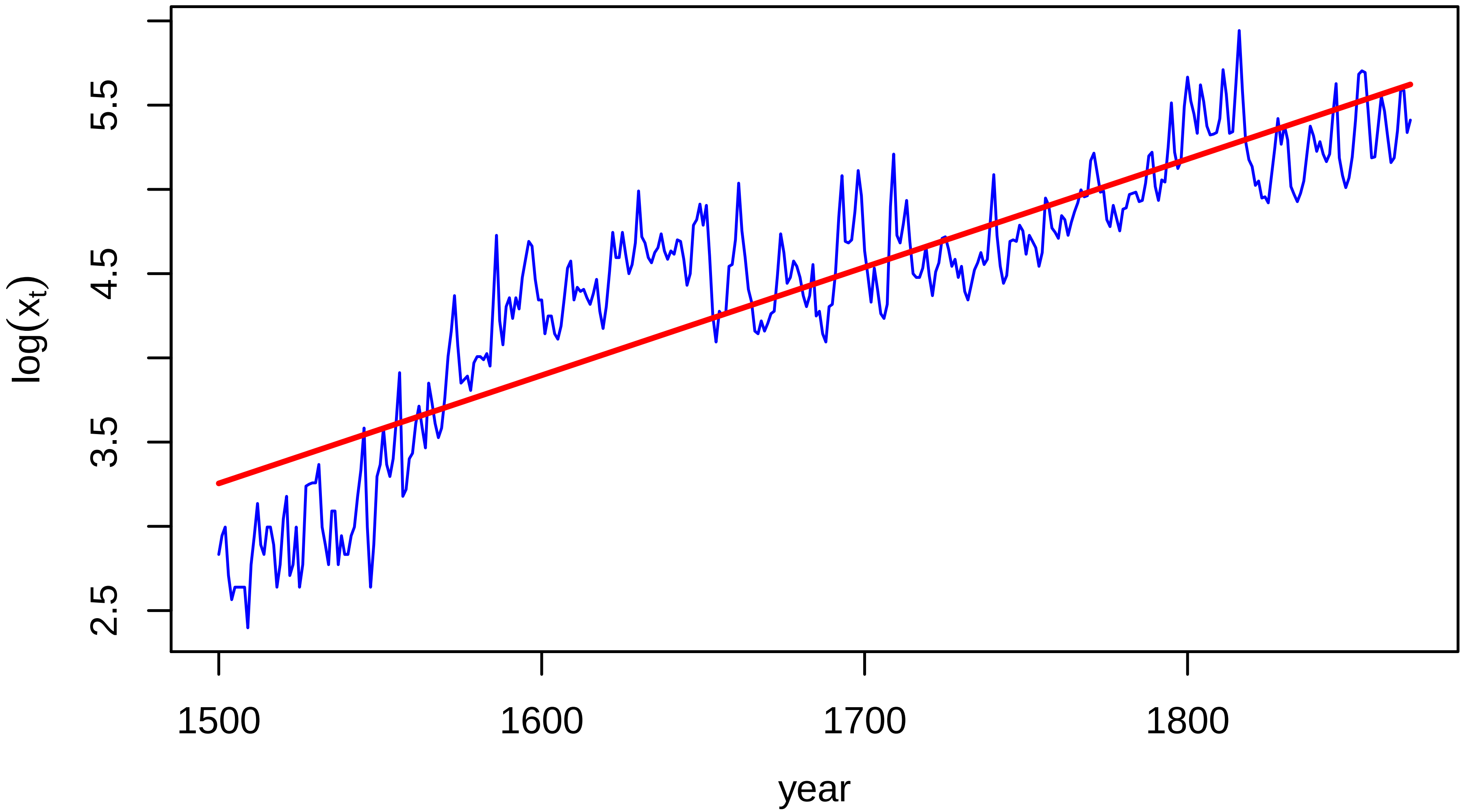
ℓ_1 **Trend Filter with** $\lambda = 397.4$



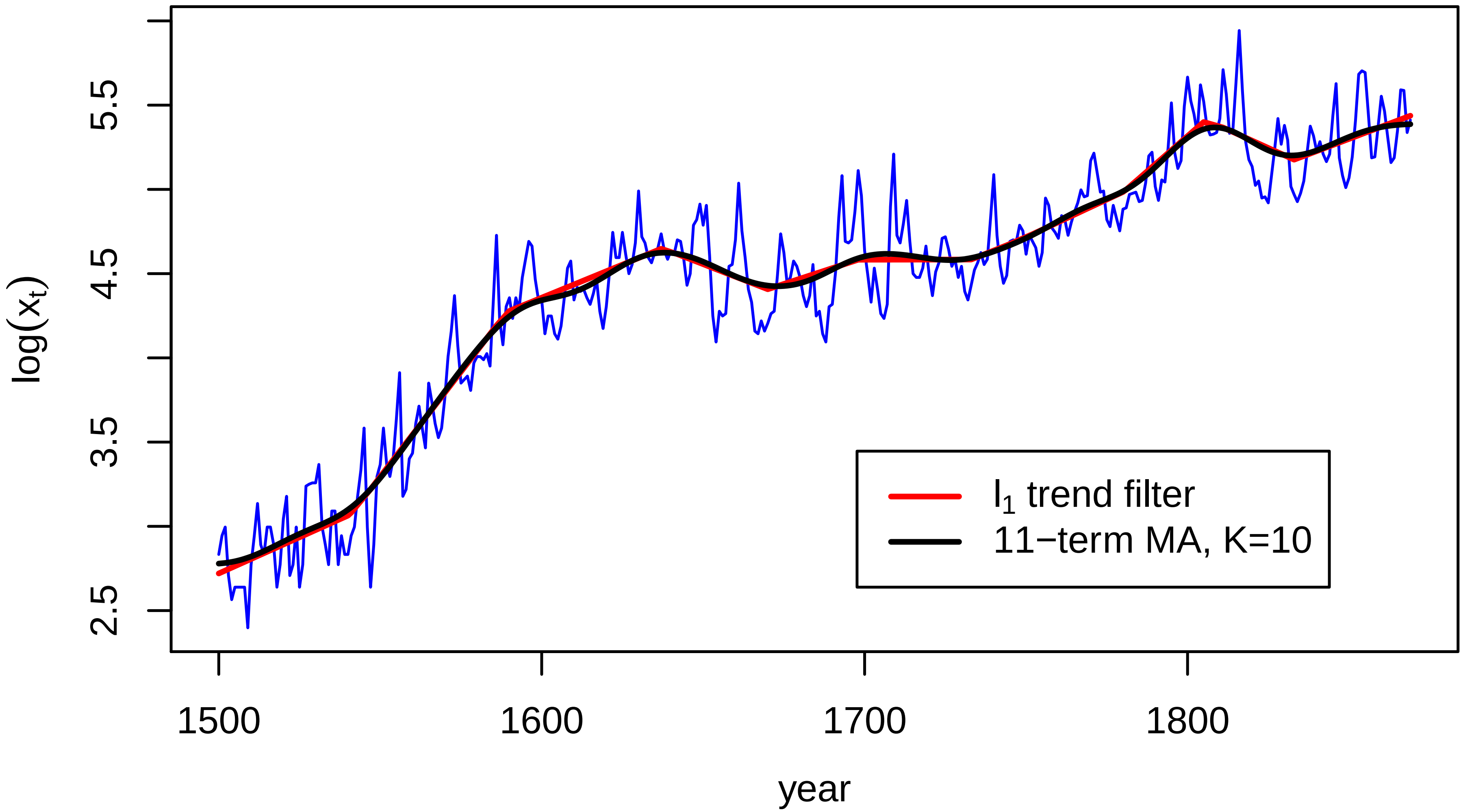


ℓ_1 Trend Filter with $\lambda = 1493.8$
log(x_t)
5.5
4.5
3.5
2.5
1500
1600
1700
1800
year

ℓ_1 Trend Filter with $\lambda = 2043.5$



ℓ_1 Trend Filter with $\lambda = 6.8$



ℓ_1 Trend Filtering: II

- interesting alternates to ℓ_1 trend filtering quantify smoothness based on something other than $\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}$
- as will be discussed shortly, $\hat{m}_{t+1} - 2\hat{m}_t + \hat{m}_{t-1}$ is a second-order differencing (analogous to a second derivative)
- replacing second-order differencing with first-order differencing $\hat{m}_t - \hat{m}_{t-1}$ (analogous to a first derivative) leads to trend estimates that are piecewise constant (rather than piecewise linear)
- replacing second-order differencing with third-order differencing $\hat{m}_{t+1} - 3\hat{m}_t + 3\hat{m}_{t-1} - \hat{m}_{t-2}$ (analogous to a third derivative) leads to trend estimates that are piecewise quadratic

ℓ_1 Trend Filtering: III

- Kim et. al (2009) and Tibshirani (2014) study properties of estimated trends for model $X_t = m_t + Y_t$ under restrictive assumption that $Y_1, \ldots, Y_n$ are independent (in the context of time series, an unappealing assumption)

A Cautionary Note on Trend Estimation: I

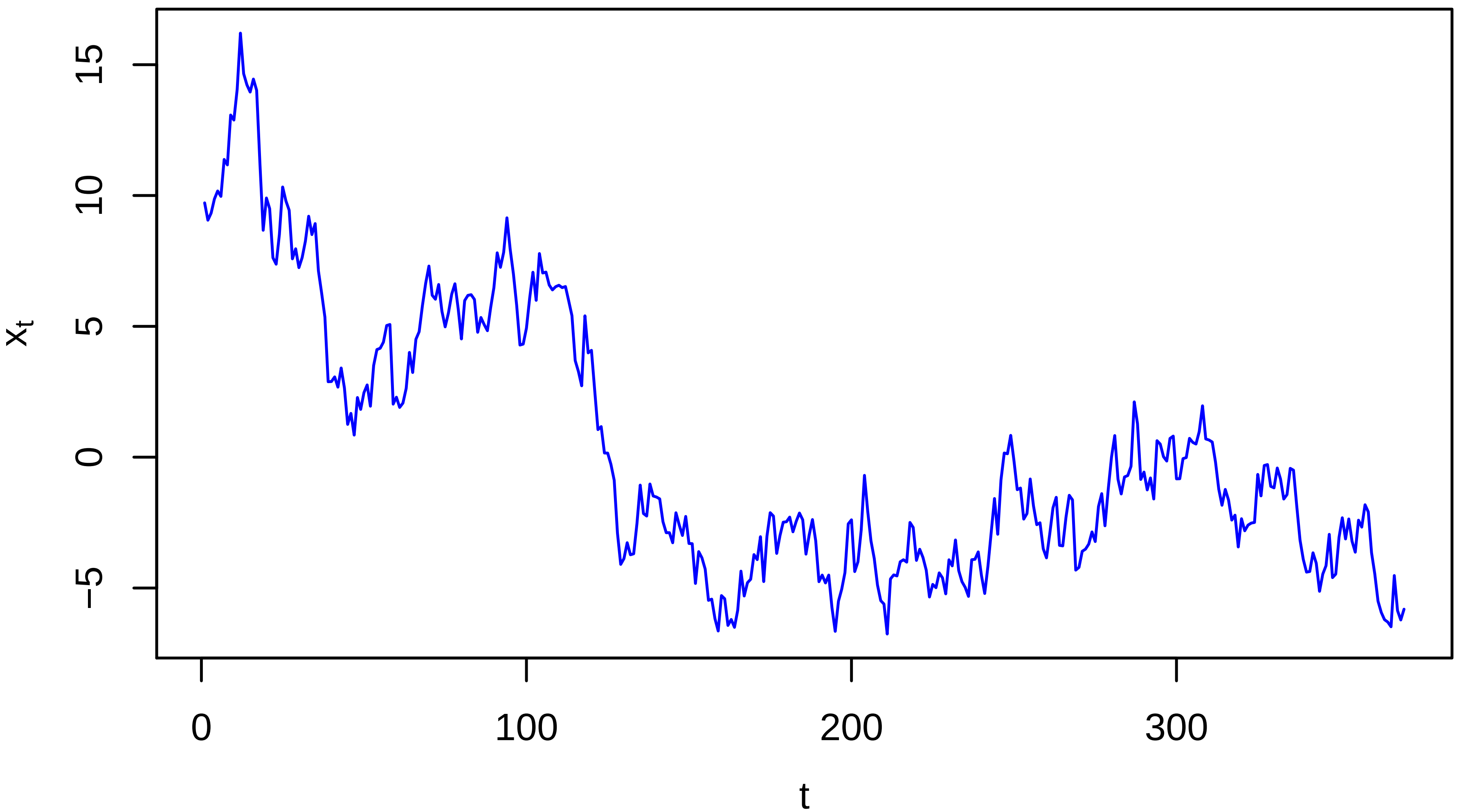
- for certain time series $\{X_t\}$, can be difficult to distinguish between models $X_t = m_t + Y_t$ and $X_t = Y_t$ (i.e., no trend), where $\{Y_t\}$ is a stationary process
- as an example, suppose that $\{Y_t\}$ is the following zero-mean AR(1) process:

$$Y_t = 0.99 Y_{t-1} + Z_t,$$

  where $\{Z_t\} \sim \text{WN}(0, 1)$ with a Gaussian distribution

  – note: close to random walk process $Y_t = Y_{t-1} + Z_t$
- suppose we set the trend m_t to zero so that $X_t = Y_t$
- next overhead shows one realization of $X_1, X_2, \ldots, X_{370}$ – same length as Beveridge wheat price index series ($n = 370$)

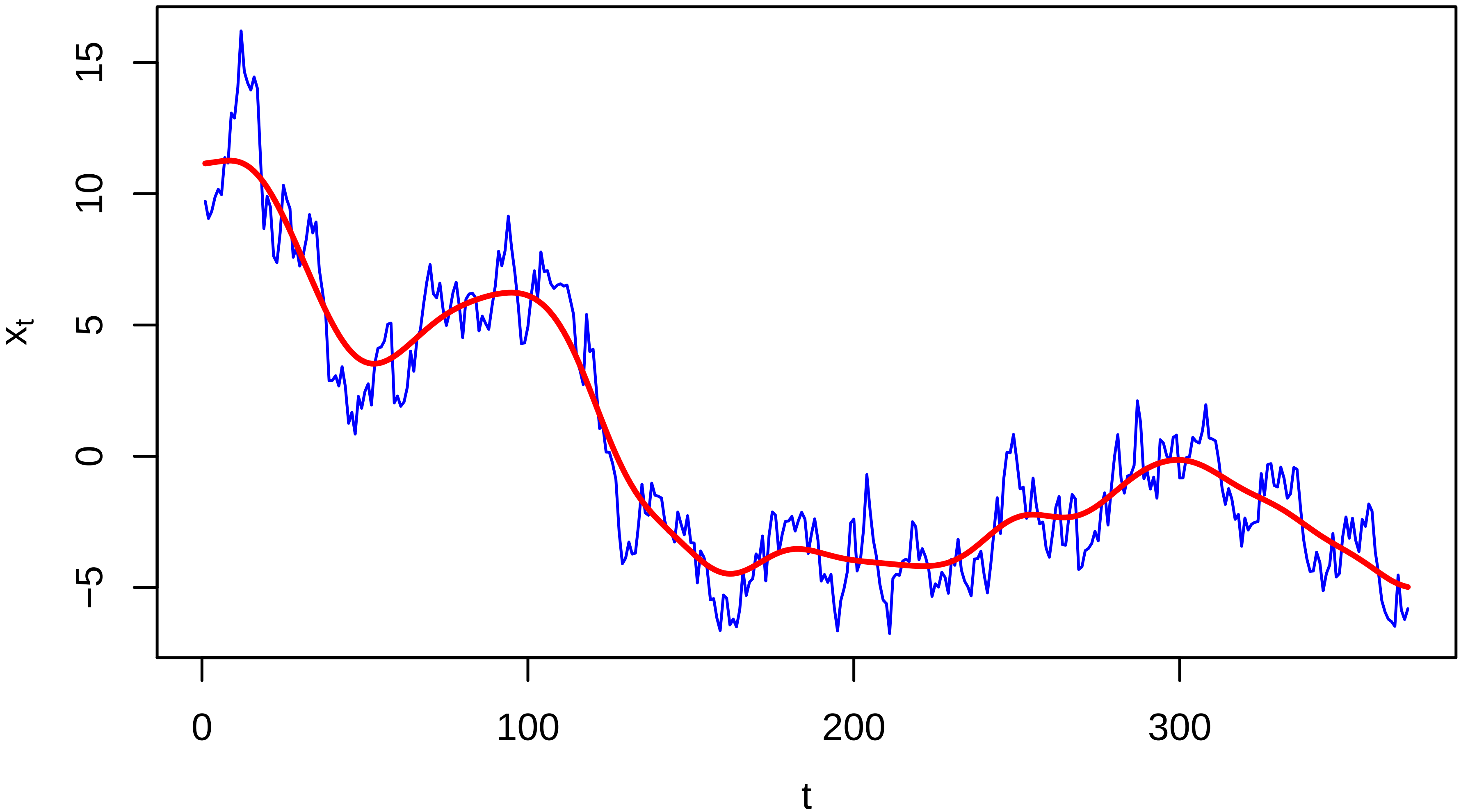
$\phi = 0.99$ AR(1) $\{x_t\}$ from Gaussian WN(0,1)



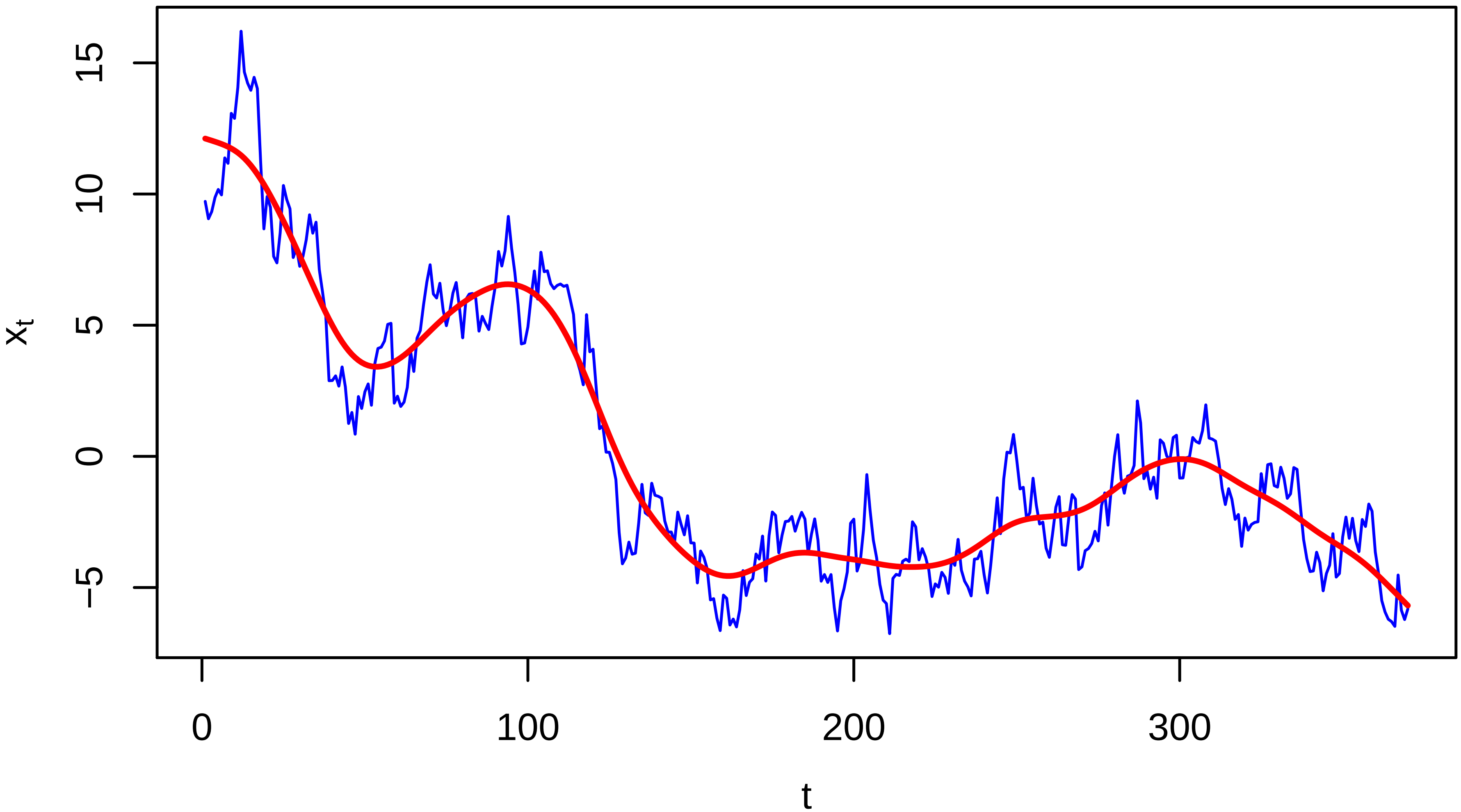
A Cautionary Note on Trend Estimation: II

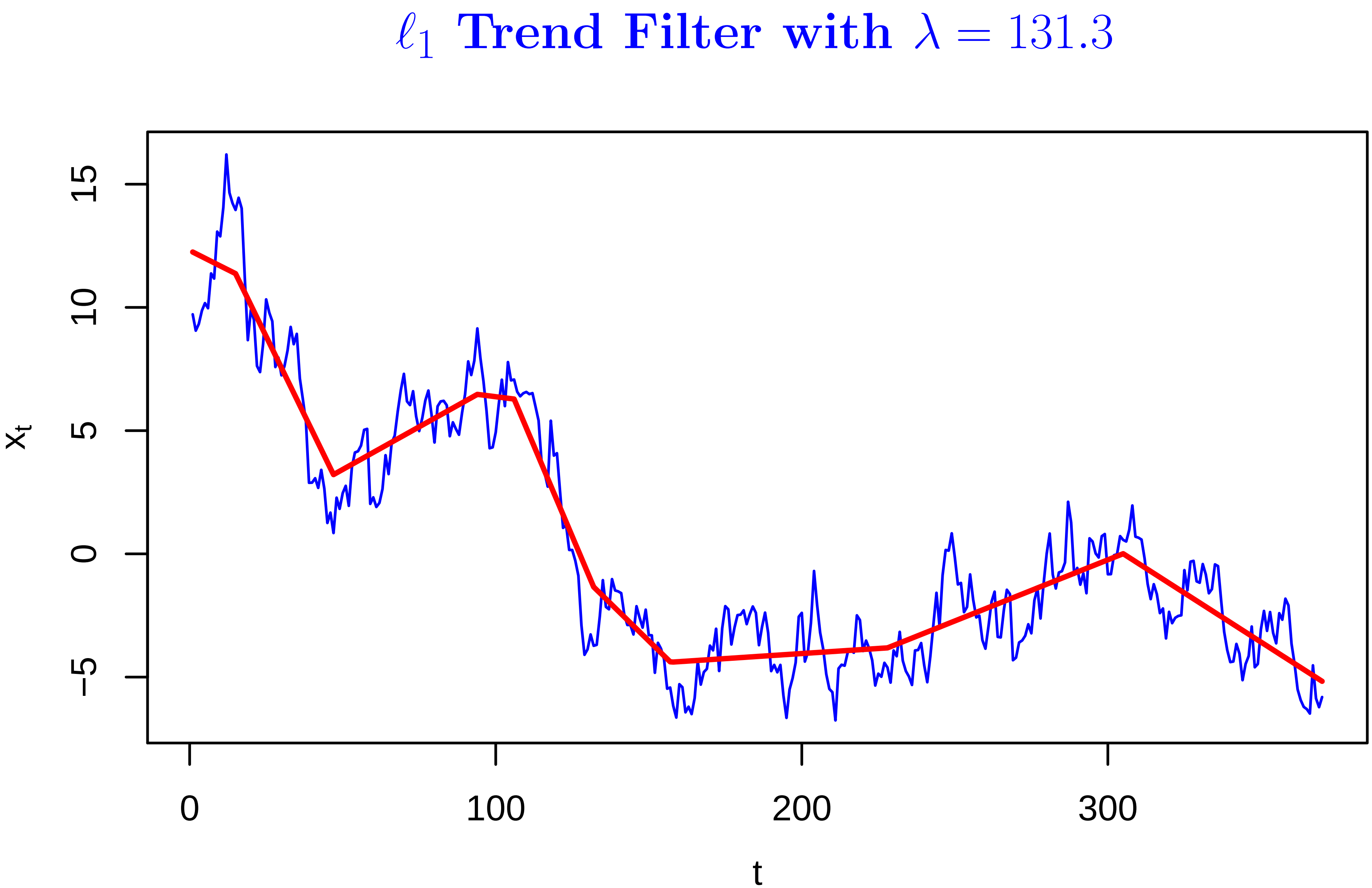
- despite absence of a nontrivial trend, model $X_t = m_t + Y_t$ superficially appears appropriate for displayed x_t
- application of any of the trend estimation procedures considered above pulls out what appears to be a nontrivial trend
- following overheads show three such trend estimates $\hat{m}_t$

$K = 10$ Applications of 11-Term MA Smoother



Hodrick–Prescott Filter with $\lambda = 8192$




ℓ_1 Trend Filter with $\lambda = 131.3$
15
10
5
0
−5
x_t
0
100
200
300
t

A Cautionary Note on Trend Estimation: III

- economic considerations say that upward trend in log of Beveridge wheat price index is reasonable – hence estimated trend $\hat{m}_t$ is arguably a reasonable descriptor of this time series
- for artificial AR(1) series, estimated trend does not have a solid basis – in fact we know there is no real trend in x_t
- to assess significance of trend component in, e.g., environmental time series, Smith (1993) advocated doing so within the context of stochastic models that can produce trend-like realizations (as in our AR(1) example) – these models can serve as null hypotheses for assessing the significance of an estimated trend component
- for details on this approach, see Craigmile et al. (2004)

Trend Elimination by Differencing: I

- focus so far has been on estimating trend m_t, which – at least in the case of polynomial fitting or ℓ_1 trend filtering – can lead to a way of forecasting trend component (examples for log of Beveridge wheat price index demonstrate that forecasts can depend quite a bit on choice of polynomial order k)
- once trend has been estimated via $\hat{m}_t$, can form residuals $r_t = x_t - \hat{m}_t$, which can be used to deduce statistical properties of stationary process $\{Y_t\}$ in model $X_t = m_t + Y_t$
- rather than estimating m_t, can take approach of eliminating it, i.e., reducing it to a constant via a *differencing operation* (presumes that $\{m_t\}$ is expressible as a low-order polynomial)

Trend Elimination by Differencing: II

- accordingly, define $B(\cdot)$ to be an operator that maps sequence $\{X_t\}$ into a new sequence $\{V_t\}$, where $V_t = X_{t-1}$ for all t:

$$B(\{X_t : t \in \mathbb{Z}\}) = \{X_{t-1} : t \in \mathbb{Z}\}$$

  (operators, filters and functionals are names for the same notion – a mapping of sequences/functions to other sequences/functions)

- $B(\cdot)$ is known as the *backward shift* operator
- above notation is too bulky, so usually simplified to just

$$BX_t = X_{t-1}$$

- define unit-lag difference operator in terms of B:

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

  (also called first-order backward difference operator)

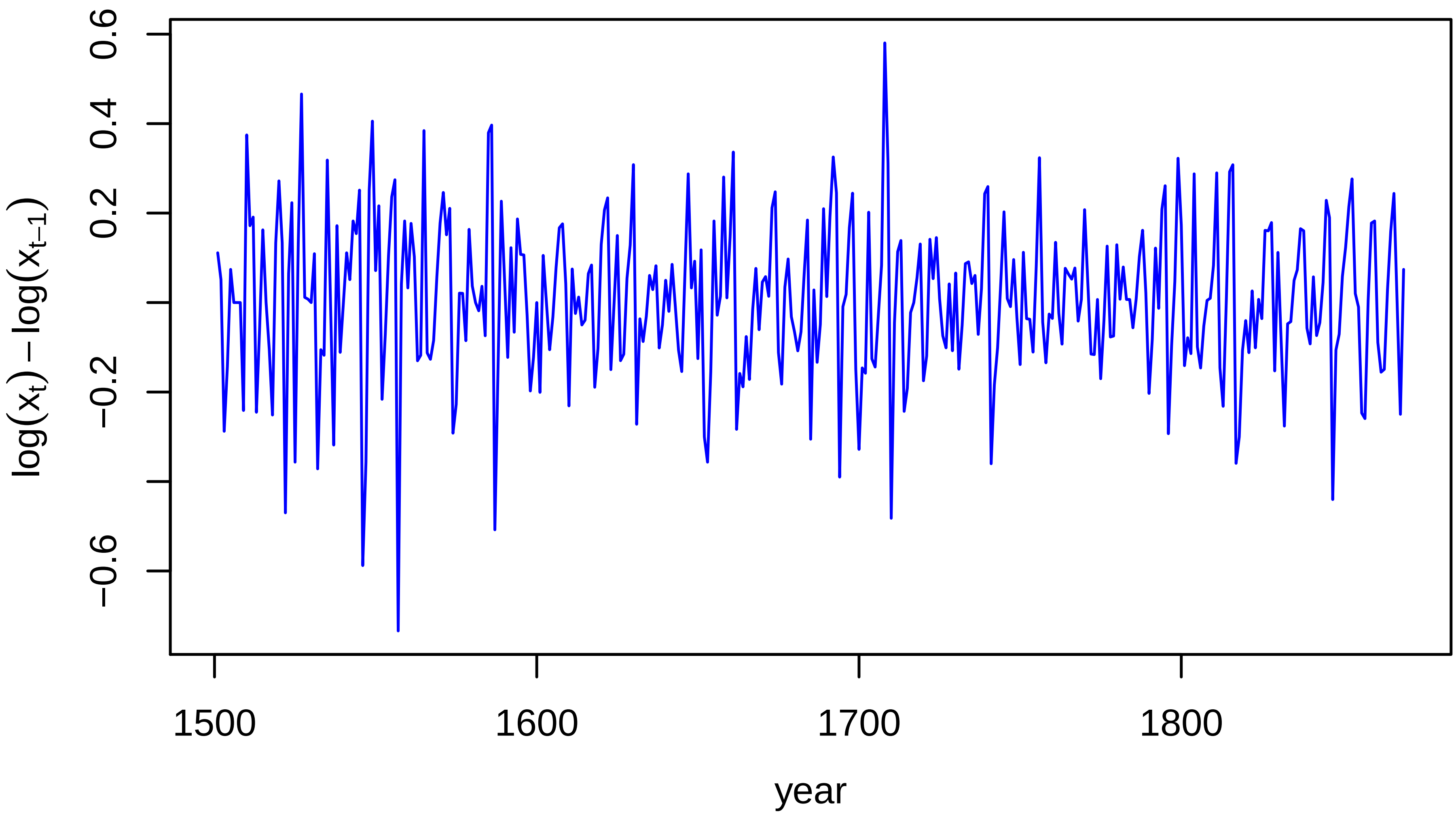
Trend Elimination by Differencing: III

- powers of B and ∇ are defined recursively: with $B^0 X_t \stackrel{\text{def}}{=} X_t$,
$$B^j X_t = B(B^{j-1} X_t) = \cdots = X_{t-j}$$
and, with $\nabla^0 X_t \stackrel{\text{def}}{=} X_t$ also,
$$\nabla^j X_t = \nabla(\nabla^{j-1} X_t)$$
for all integers $j \geq 1$
- for example,
$$\begin{aligned}\nabla^2 X_t &= \nabla(\nabla X_t) \\ &= (1 - B)(1 - B) X_t \\ &= (1 - 2B + B^2) X_t \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$
defines the second-order backward difference operator

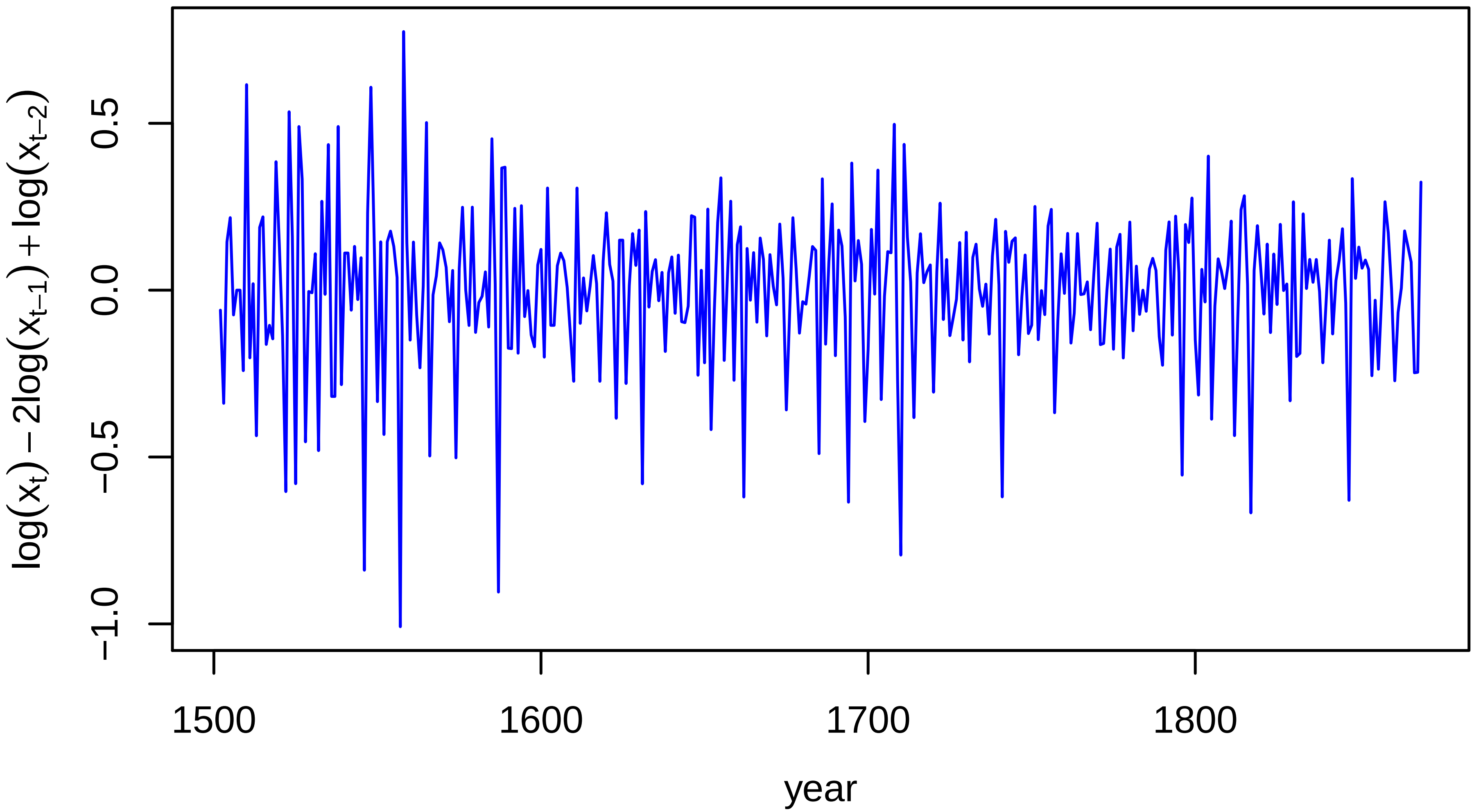
Trend Elimination by Differencing: IV

- suppose $\{m_t\}$ is a linear trend: $m_t = c_0 + c_1 t$
- application of first-order backward difference operator yields
$$\nabla m_t = m_t - m_{t-1} = c_0 + c_1 t - (c_0 + c_1(t-1)) = c_1;$$
i.e., operator ∇ reduces linear trend to a constant
- exercise: a polynomial trend of degree k can be reduced to a constant by application of ∇^k
- another exercise: if $\{Y_t\}$ is a stationary process, then so is $\{\nabla^k Y_t\}$ for any k
- hence, if $X_t = m_t + Y_t$, where $m_t = c_0 + \cdots + c_k t^k$ is a kth order polynomial and $\{Y_t\}$ is a stationary process, then $\nabla^k X_t = \nabla^k m_t + \nabla^k Y_t$ is a stationary process with nonzero mean (in fact the mean is given by $k!c_k$)

1st Difference of Log(Beveridge Wheat Price Index)



2nd Difference of Log(Beveridge Wheat Price Index)



Trend & Seasonal Estimation and Elimination: II

- methods for estimating and eliminating m_t in model $X_t = m_t + Y_t$ can be extended to handle both a trend and a seasonal component in classical decomposition model

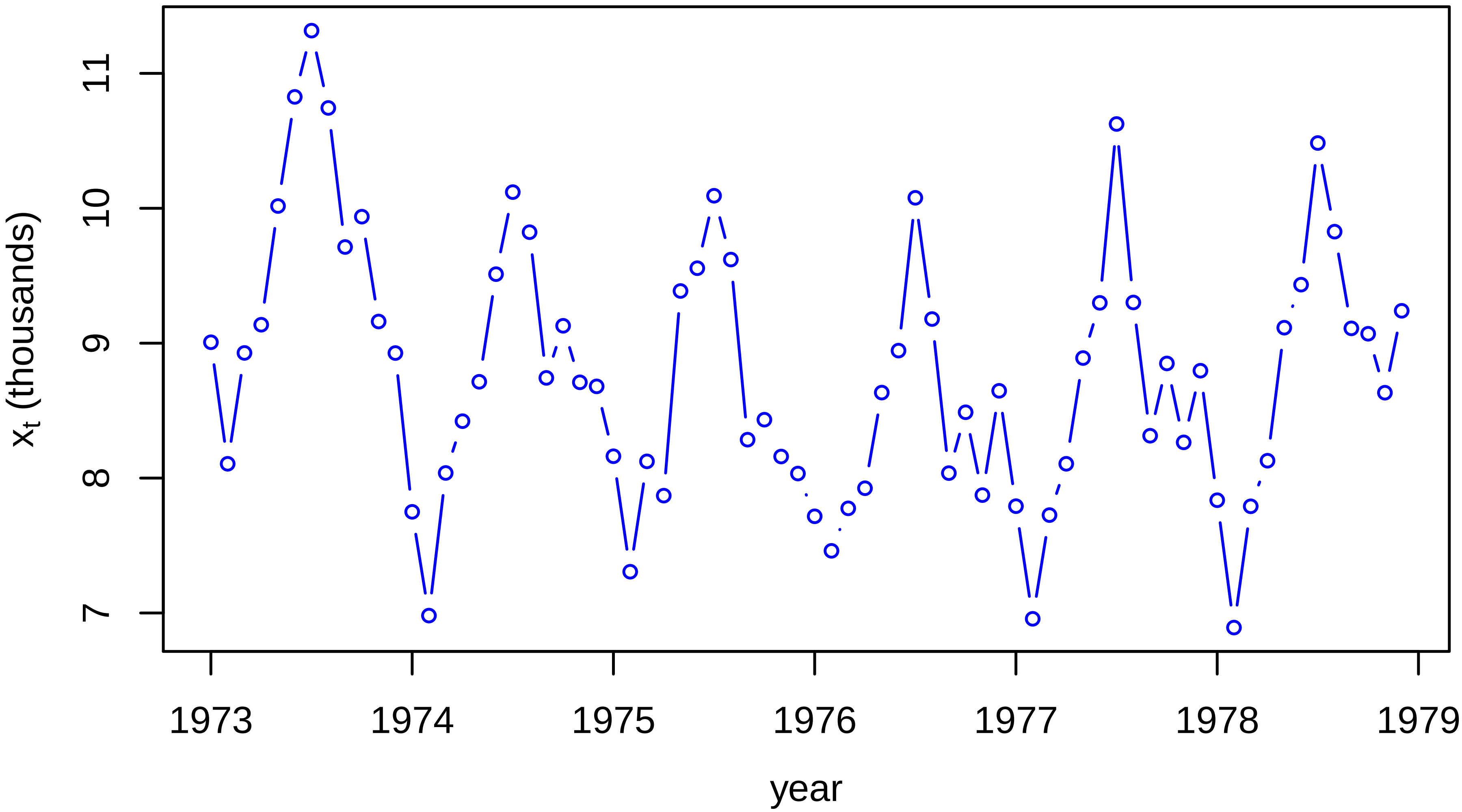
$$X_t = m_t + s_t + Y_t, \quad t = 1, \ldots, n,$$

where we recall that $\{s_t\}$ has a known period d (i.e., $s_{t+d} = s_t$ for all t), and we assume that

$$\sum_{j=1}^{d} s_{t+j} = 0 \text{ for all } t$$

- to illustrate methodology, let's look at monthly time series of accidental deaths (data from Brockwell & Davis)

Monthly Counts of Accidental Deaths in USA



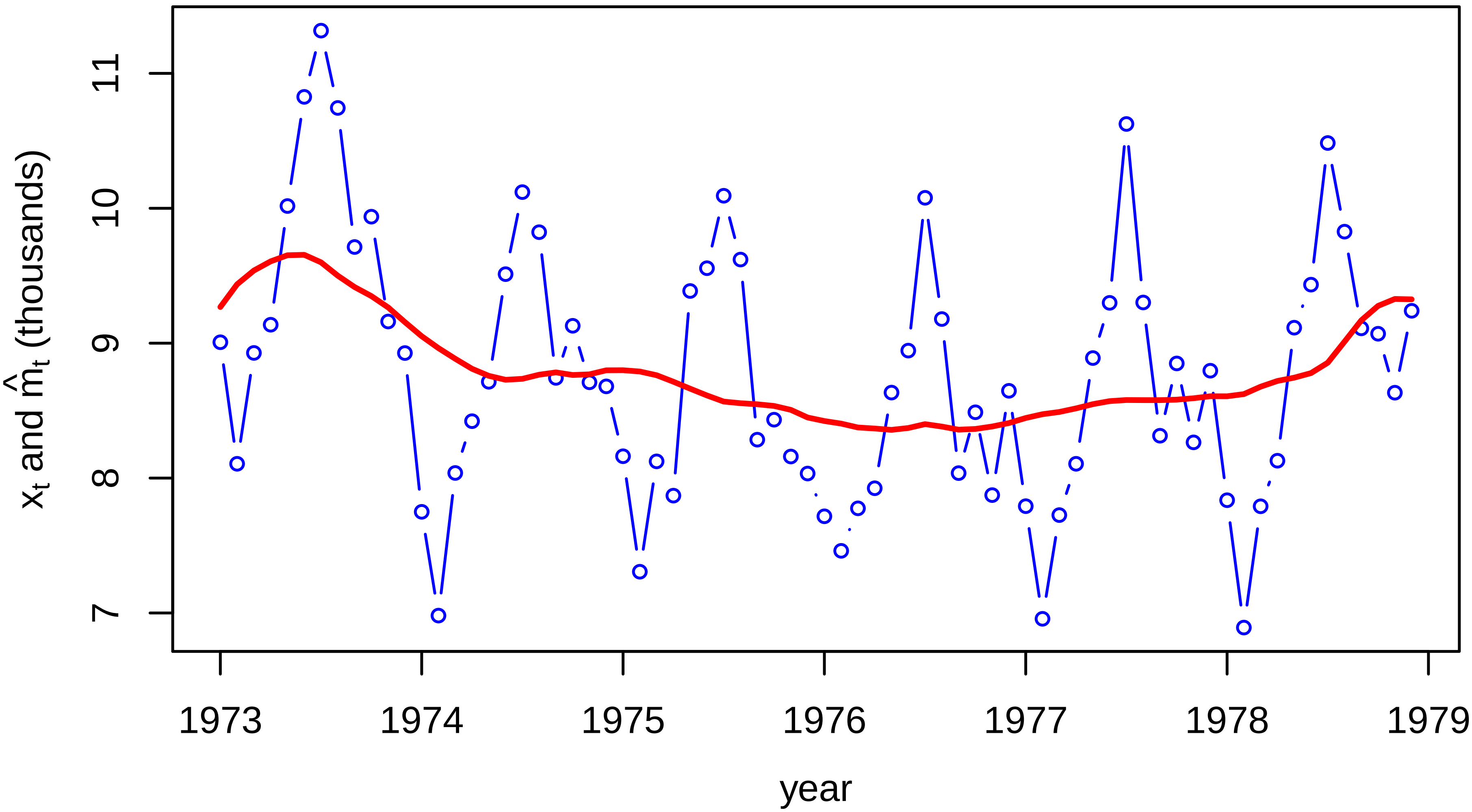
Trend & Seasonal Estimation: I

- first step is to get preliminary estimate of trend $\{m_t\}$ using a smoothing filter that eliminates seasonal component $\{s_t\}$
- obvious choice is a 12-term moving average smoother of length $d = 12$ since $\sum_{j=1}^{d} s_{t+j} = 0$; however, to avoid undesirable time shifts, need to use a two-sided moving average of odd length $2q+1$, which conflicts with $d = 12$ (bummer!)
- as an alternative, use 13-term two-sided moving average smoother $\{a_j\}$, but set a_{-6} and a_6 to half the values of other a_j's:

$$\{a_j : j = -6, \ldots, 6\} = \left\{\frac{1}{24}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{24}\right\}$$

- note: $\sum_j a_j = 1$, as is required for smoothing filters
- note: both $\frac{1}{24}$ weights are applied to same month – thus 13- and 12-term smoothers both eliminate seasonal component

Monthly Counts of Accidental Deaths in USA



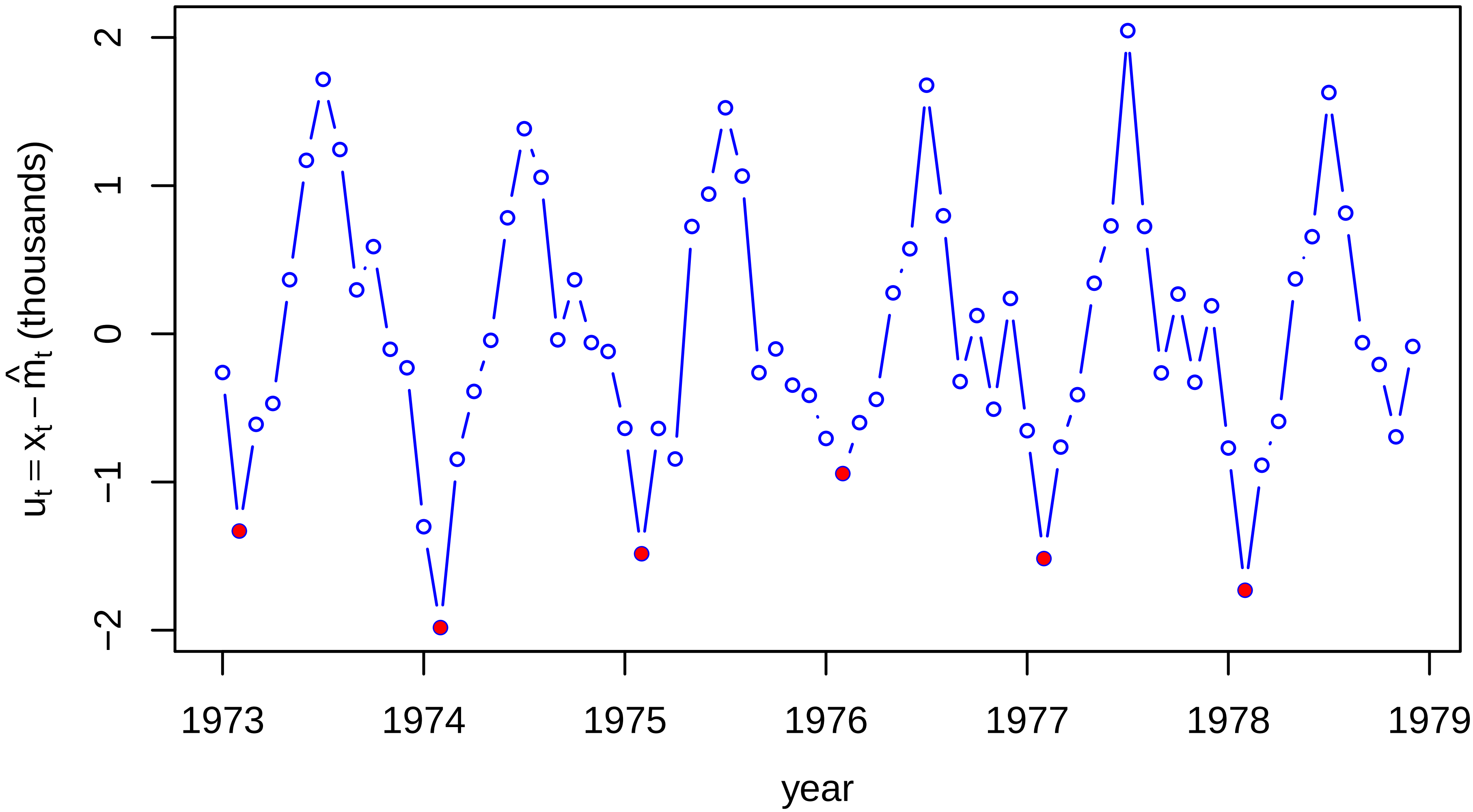
Trend & Seasonal Estimation: II

- to estimate seasonal pattern $\{s_j : j = 1, \ldots, d\}$, form

$$u_t \stackrel{\text{def}}{=} x_t - \hat{m}_t,$$

where $\{\hat{m}_t\}$ is preliminary trend estimate

Preliminary Detrending of Accidental Deaths Series



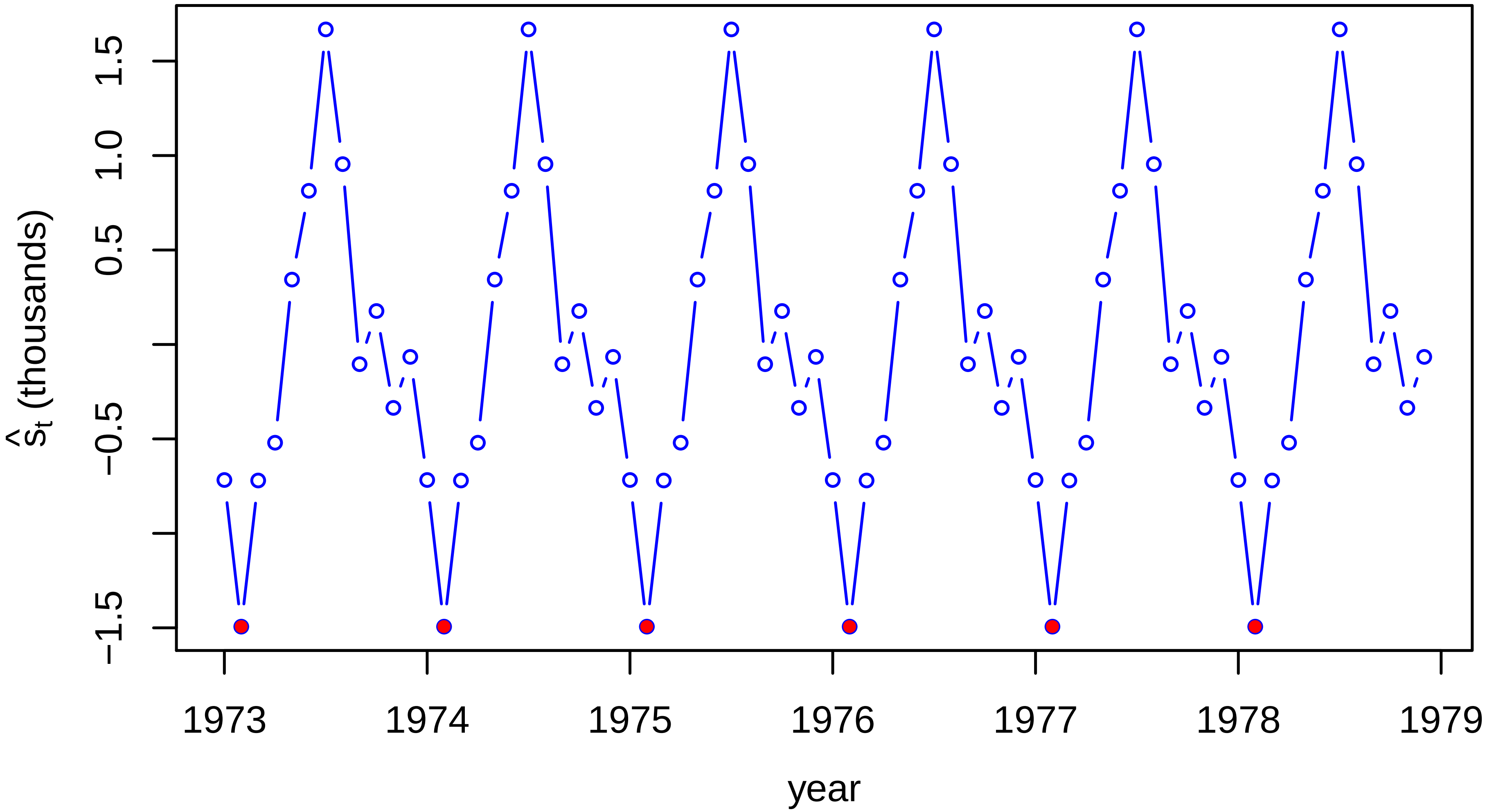
Trend & Seasonal Estimation: III

- estimate s_1 by averaging all u_t's associated with January; s_2 by averaging all u_t's associated with February; …; s_{12} by averaging all u_t's associated with December
- denote these estimates by $\{w_j : j = 1, \ldots, d\}$
- for $j = 1, \ldots, d$, estimate seasonal pattern by

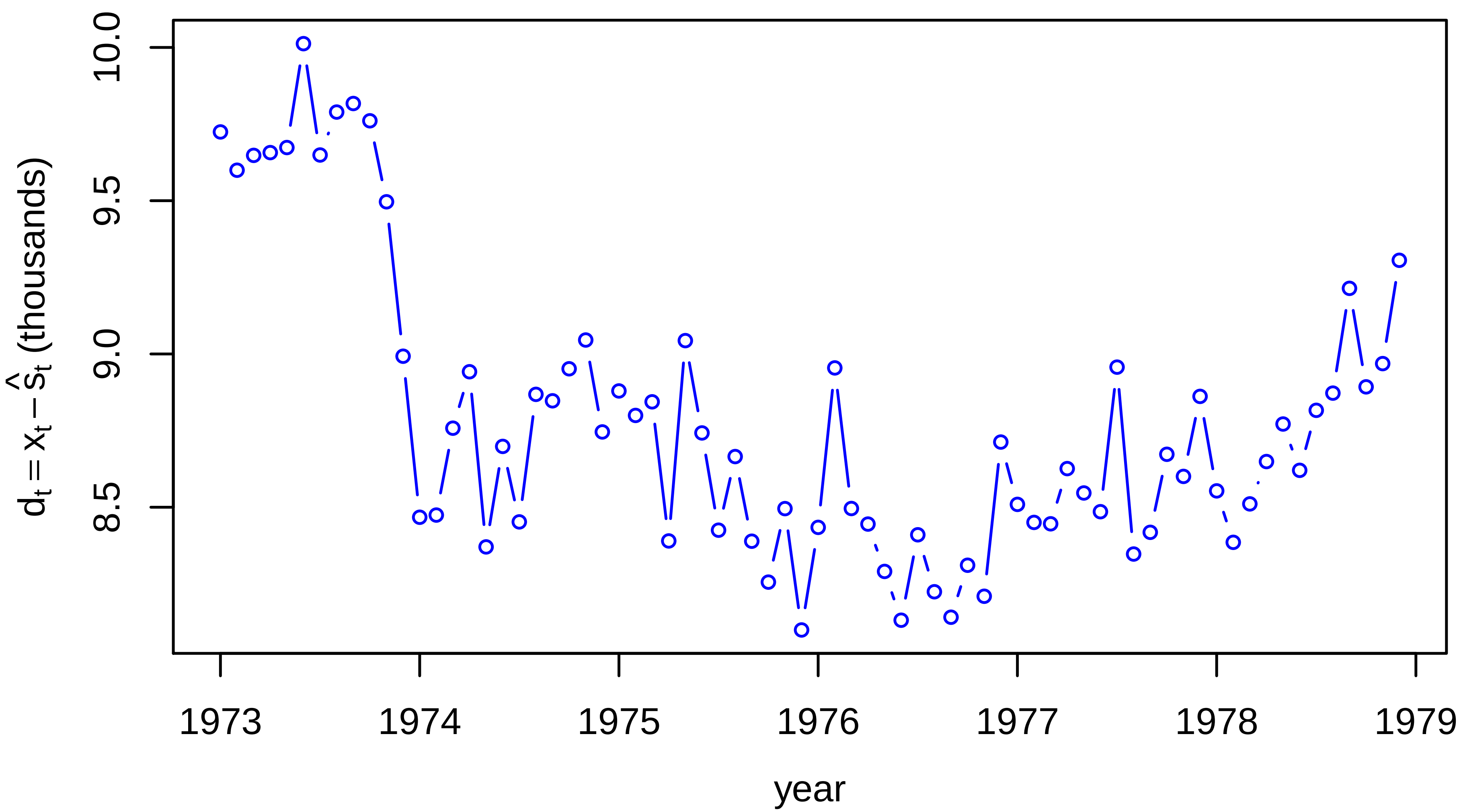
$$\hat{s}_j = w_j - \bar{w}, \text{ where } \bar{w} \stackrel{\text{def}}{=} \frac{1}{d} \sum_{j=1}^{d} w_j$$

- note: $\sum_{j=1}^{d} \hat{s}_j = 0$ mimics modeling assumption $\sum_{j=1}^{d} s_j = 0$
- to estimate $\{s_t\}$ by, say, $\{\hat{s}_t\}$, replicate estimated seasonal pattern $\{\hat{s}_j\}$ as needed – use $\{\hat{s}_t\}$ so determined to deseasonalize time series via $x_t - \hat{s}_t \stackrel{\text{def}}{=} d_t$

Estimated Seasonal Component $\{\hat{s}_t\}$ for $\{x_t\}$



Deseasonalized Data $d_t = x_t - \hat{s}_t$



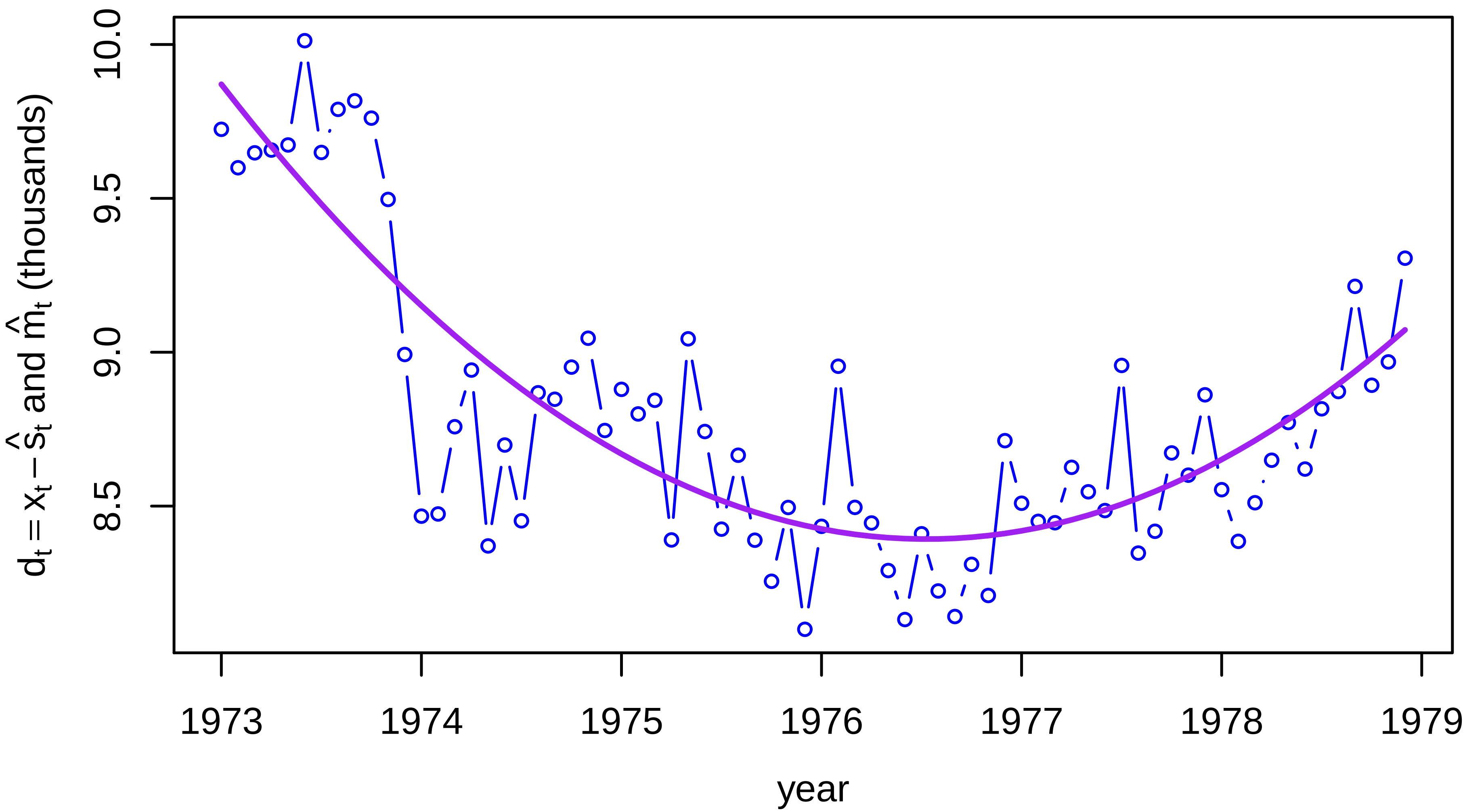
Trend & Seasonal Estimation: IV

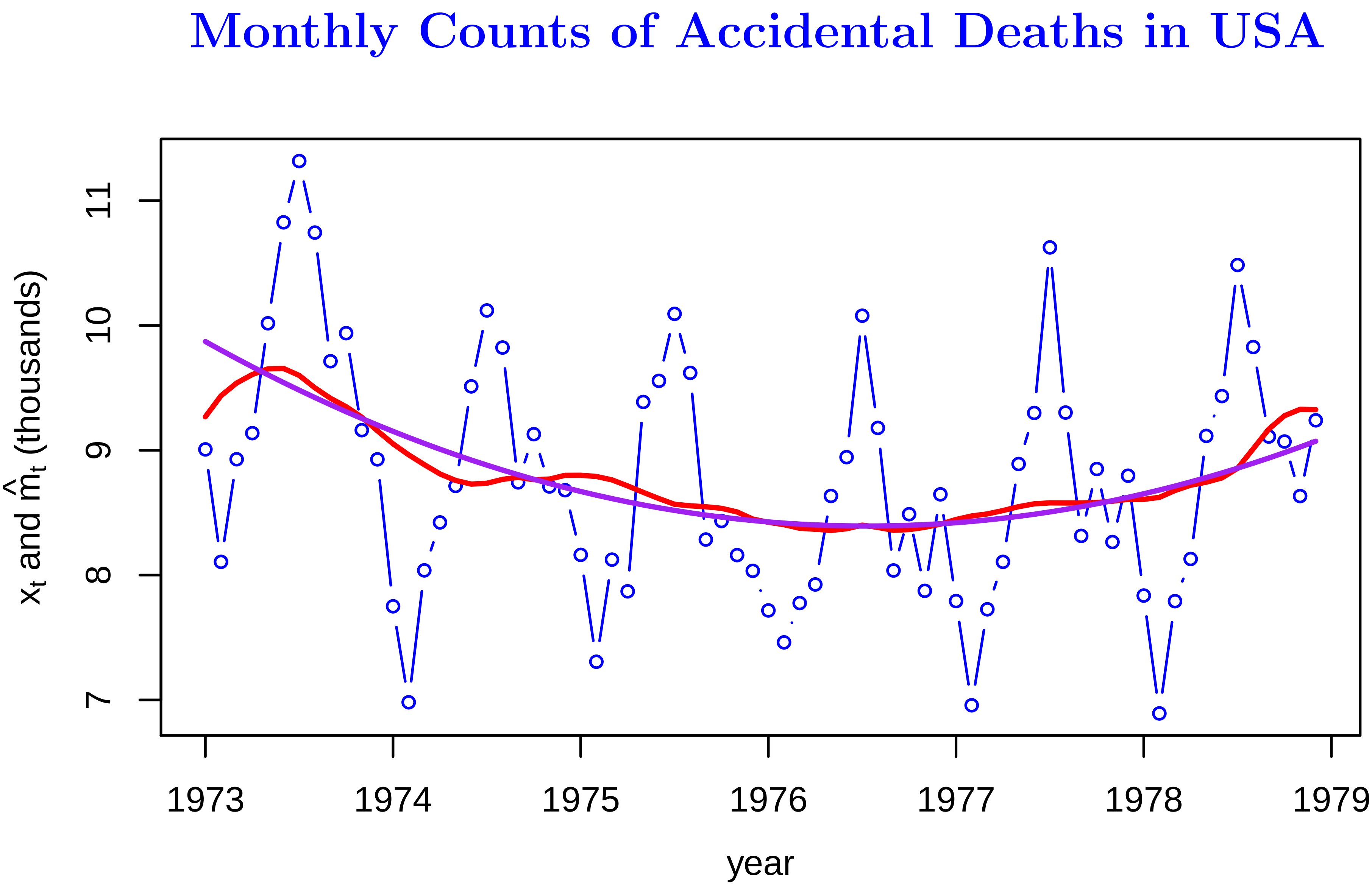
- can now reestimate trend $\{m_t\}$ using deseasonalized data $\{d_t\}$
- given the appearance of $\{d_t\}$, use of a quadratic polynomial to estimate $\{m_t\}$ seems appropriate (and extrapolation is straightforward, but might only be reasonable in the short term)
- letting $\{\hat{m}_t\}$ now denote the final trend estimate, can form residuals

$$r_t = d_t - \hat{m}_t = x_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \ldots, n,$$

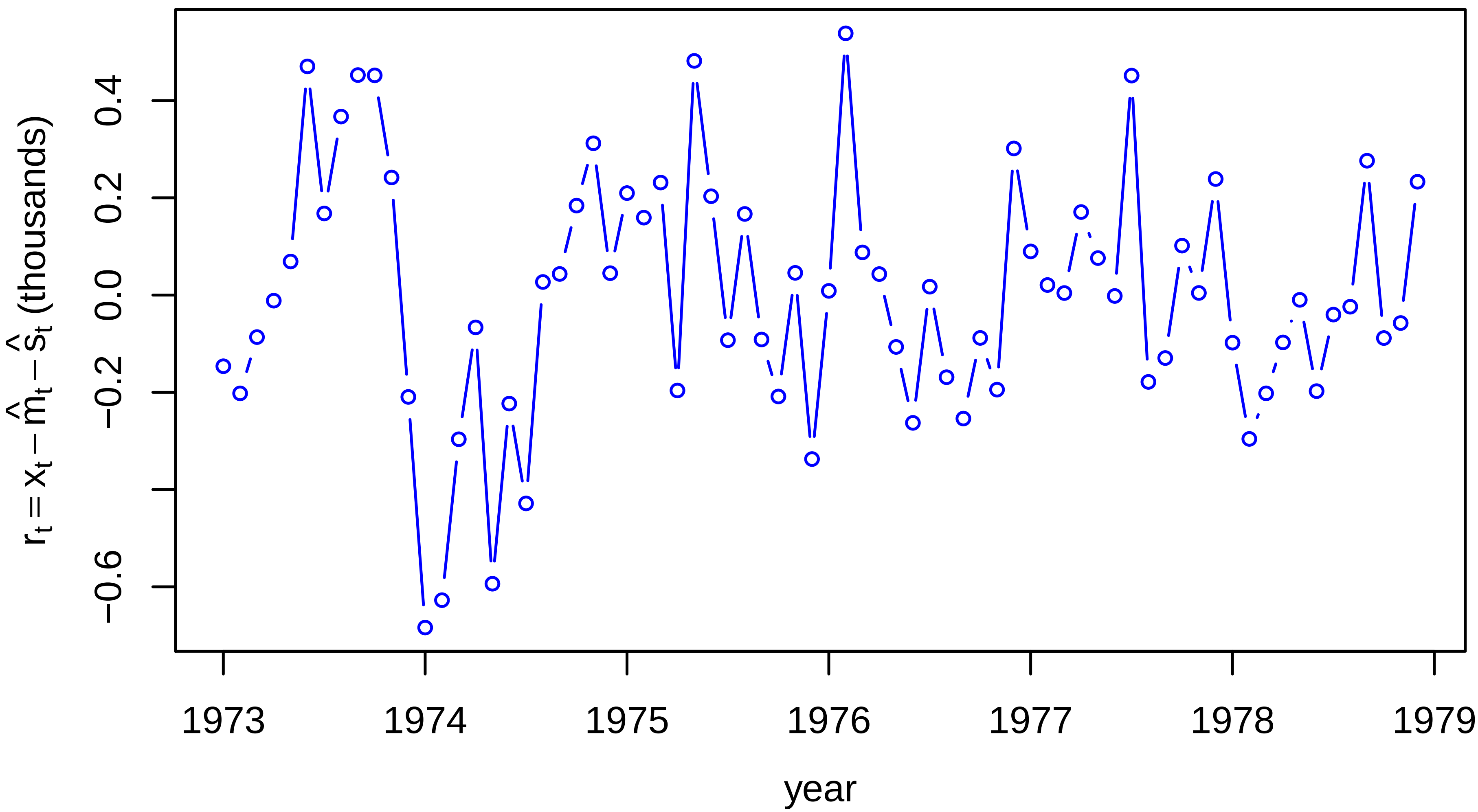
where $\{r_t\}$ is taken to be a surrogate for a realization of $\{Y_t\}$ in the model $X_t = m_t + s_t + Y_t$

Deseasonalized Data $\{d_t\}$ and Trend Estimate $\{\hat{m}_t\}$




Monthly Counts of Accidental Deaths in USA
x_t and $\hat{m}_t$ (thousands)
7
8
9
10
11
1973
1974
1975
1976
1977
1978
1979
year

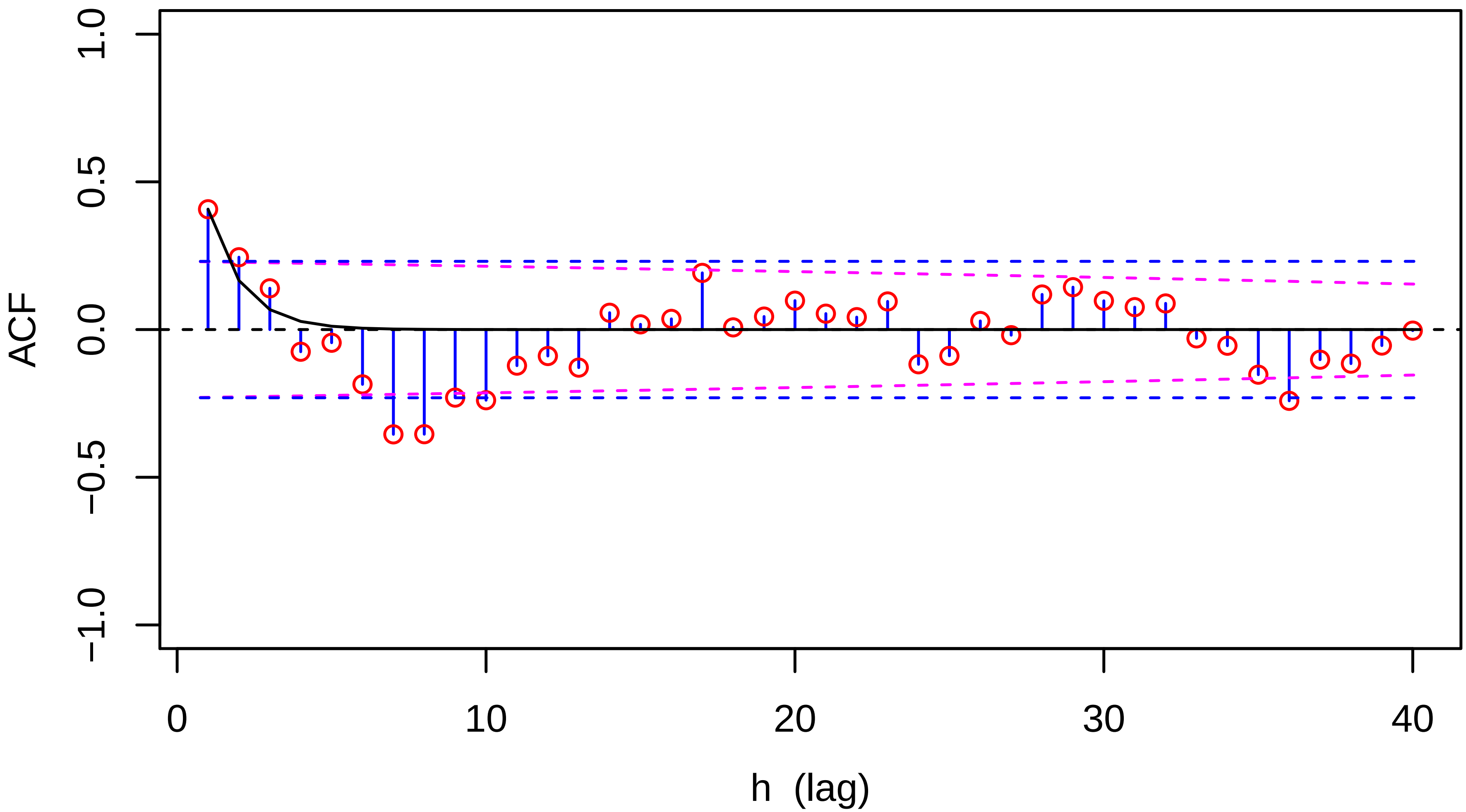
Residuals $\{r_t\}$ from Removal of $\{\hat{m}_t\}$ and $\{\hat{s}_t\}$



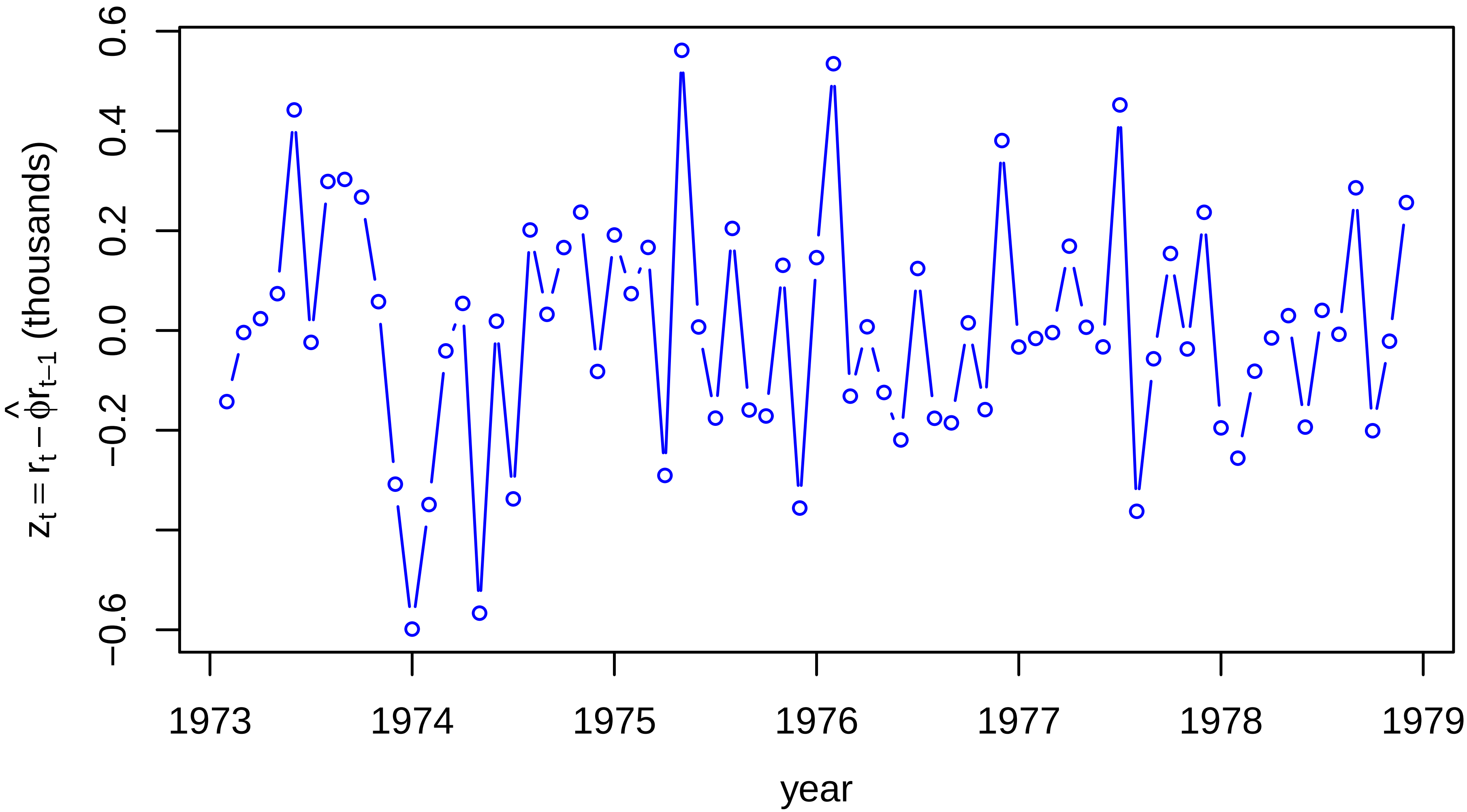
Trend & Seasonal Estimation: V

- sample ACF for $\{r_t\}$ (next overhead) suggests that a suitable model for the residuals might be an AR(1) process
- estimate ϕ for this process from sample ACF at unit lag, which yields $\hat{\phi} \doteq 0.41$
- if AR(1) model is viable, then $z_t = r_t - \hat{\phi} r_{t-1}$, $t = 2, 3, \ldots, n$, should resemble white noise, so need to look at sample ACF for $\{z_t\}$ also

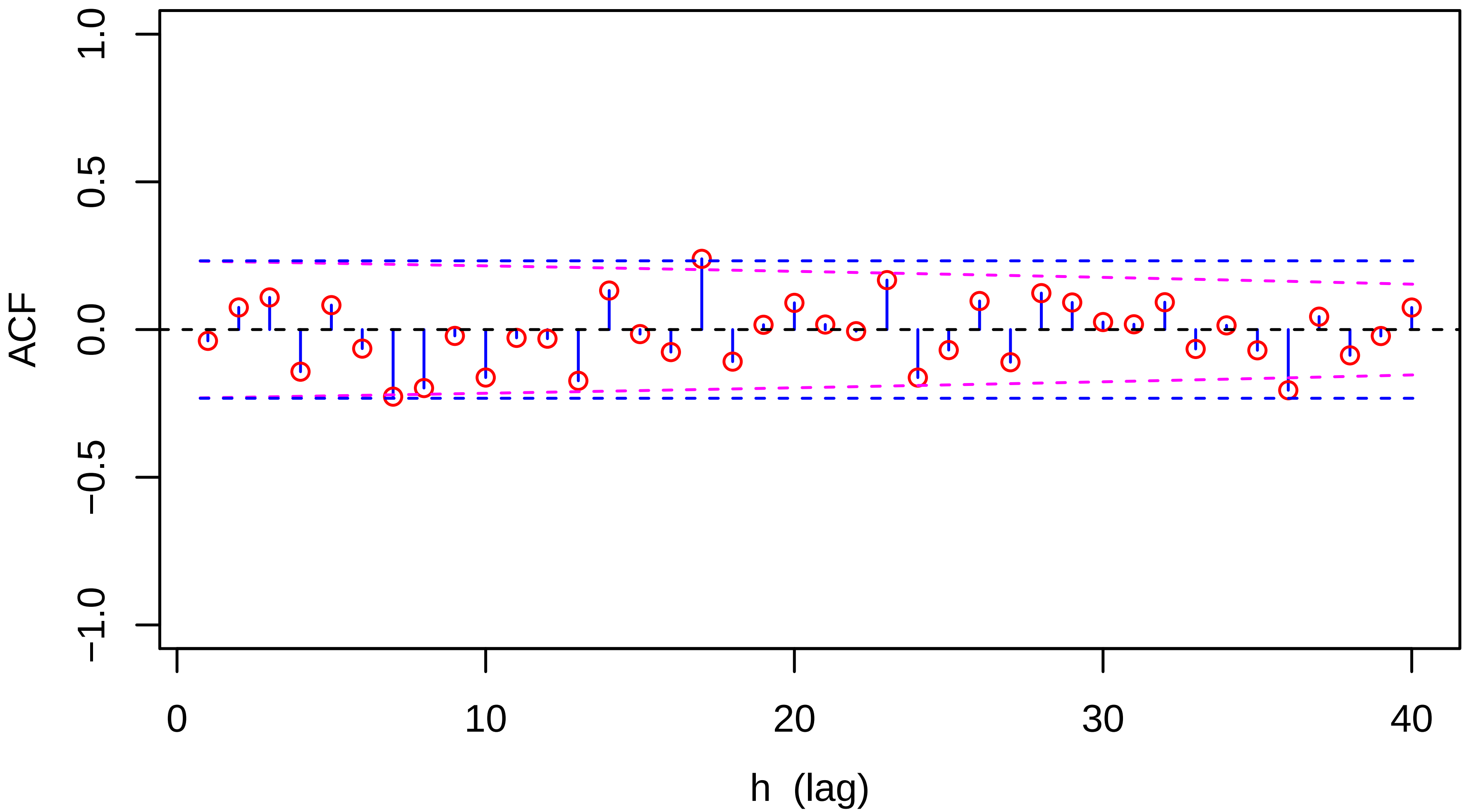
Sample ACF for $\{r_t\}$



Residuals $z_t = r_t - \hat{\phi} r_{t-1}$ from Fitted AR(1) Model



Sample ACF for $\{z_t\}$



Trend & Seasonal Estimation: VI

- now have estimates of trend and seasonal components and a viable model for stationary process $\{Y_t\}$
- let's review steps in simple procedure for estimating $\{m_t\}$ and $\{s_t\}$ in classical decomposition model

$$X_t = m_t + s_t + Y_t, \quad t = 1, \ldots, n,$$

where $\{s_t\}$ is periodic with period d and $\sum_{j=1}^{d} s_j = 0$

Trend & Seasonal Estimation: VII

1. form *preliminary* estimate $\{\hat{m}_t\}$ of trend by passing data through filter that eliminates $\{s_t\}$ as much as possible
2. subtract trend estimate from data: $u_t = x_t - \hat{m}_t$
3. seasonal pattern estimate $\{\hat{s}_j : j = 1, \ldots, d\}$ obtained by averaging u_t's for each seasonal component (denote averages by $\{w_j\}$) and then centering them:

$$\hat{s}_j = w_j - \bar{w}, \text{ where } \bar{w} \stackrel{\text{def}}{=} \frac{1}{d}\sum_{j=1}^{d} w_j$$

4. replicate $\{\hat{s}_j\}$ as need be to form estimate $\{\hat{s}_t\}$ of $\{s_t\}$
5. form deseasonalized data: $d_t = x_t - \hat{s}_t$
6. use deseasonalized data to get *final* estimate $\{\hat{m}_t\}$ of trend

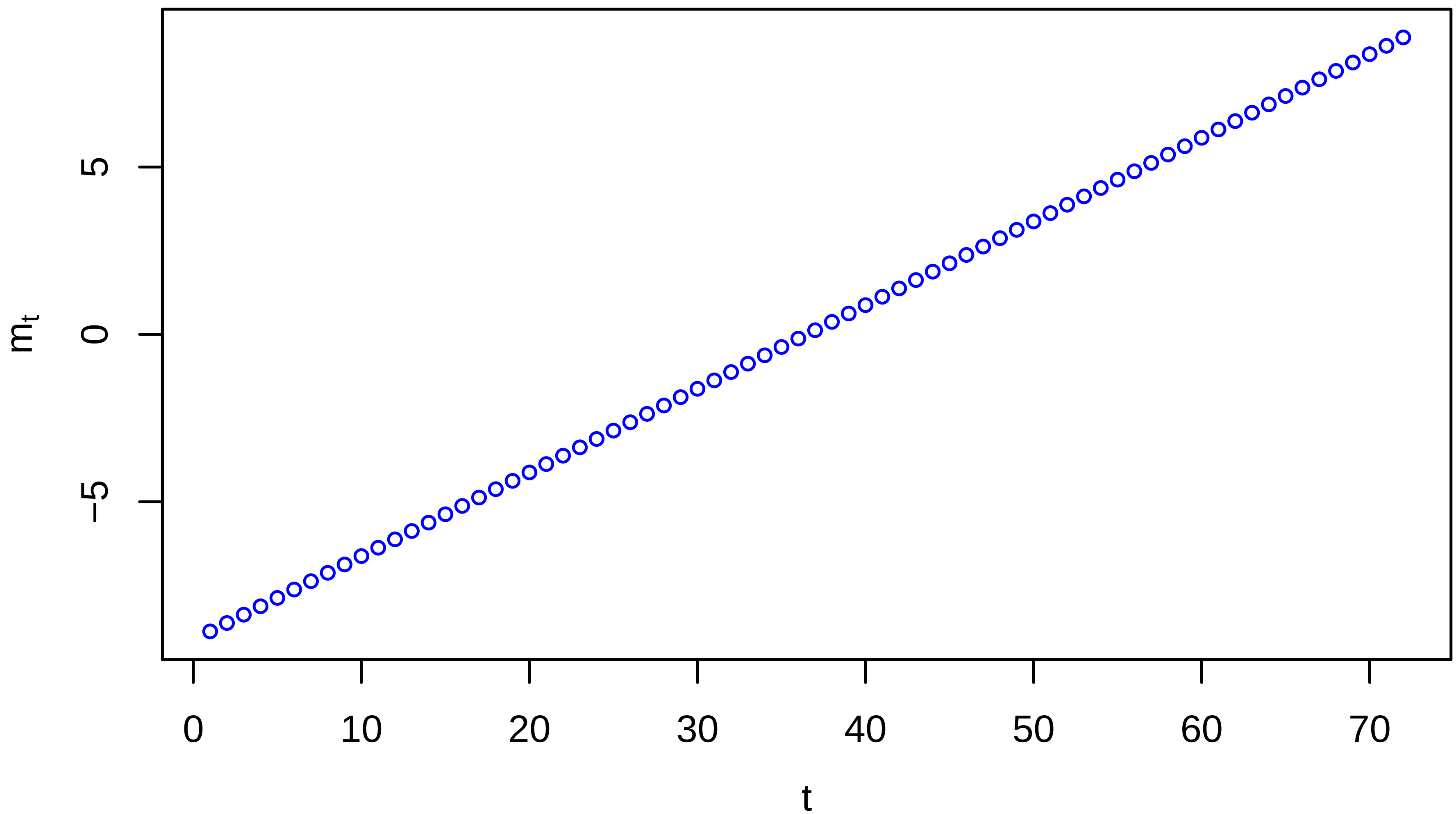
Trend & Seasonal Estimation: VIII

- Q: do we really need to do preliminary detrending?
- A: in general, yes, as following toy example demonstrates
- suppose time series is given by

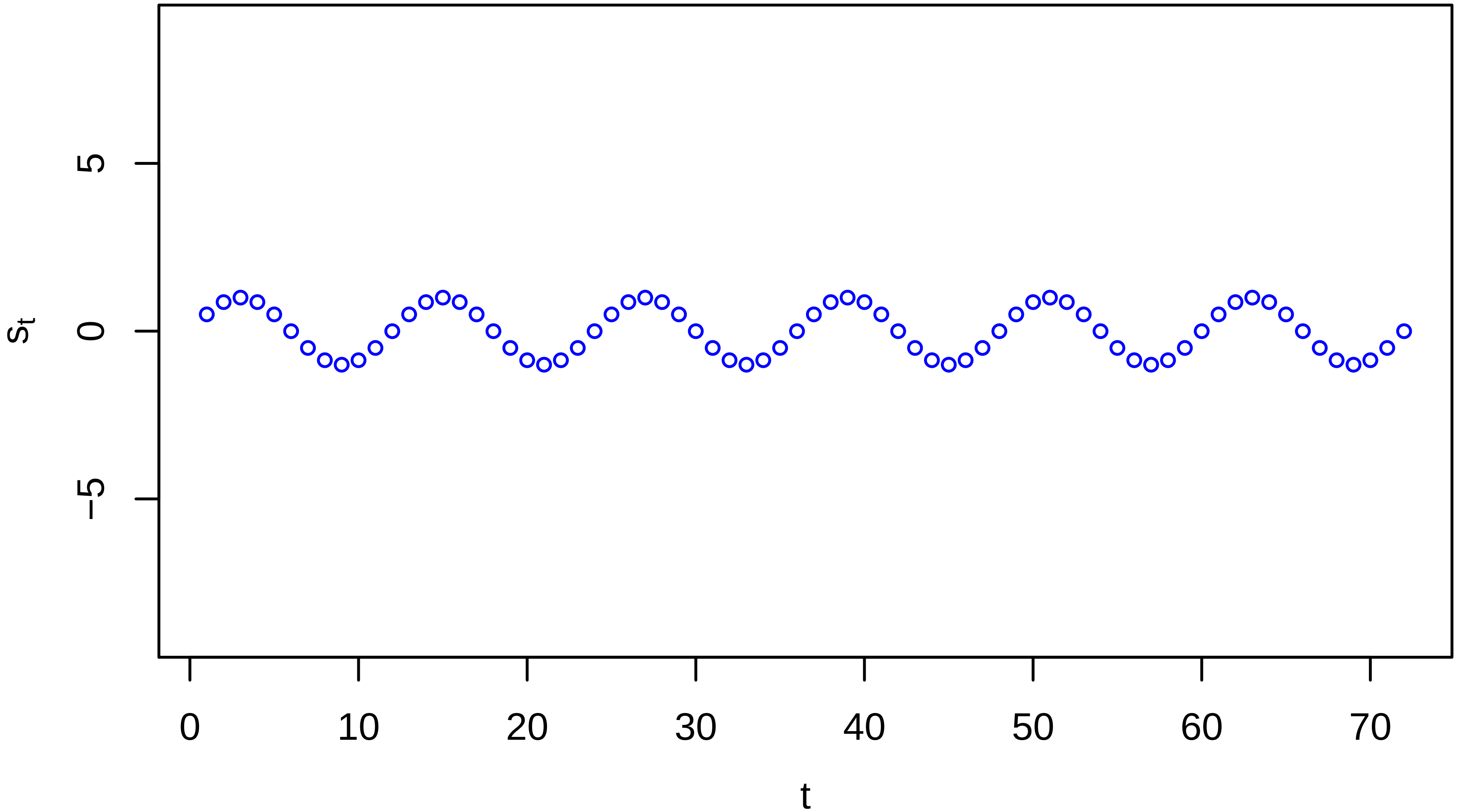
$$x_t = m_t + s_t, \quad t = 1, \ldots, 72,$$

  where $m_t = (t - 36.5)/4$ and $s_t = \sin(2\pi t/12)$ so that $\{s_t\}$ is periodic with a period of $d = 12$ (i.e., *no* stochastic noise!)
- first let's see what the recommended procedure gives us

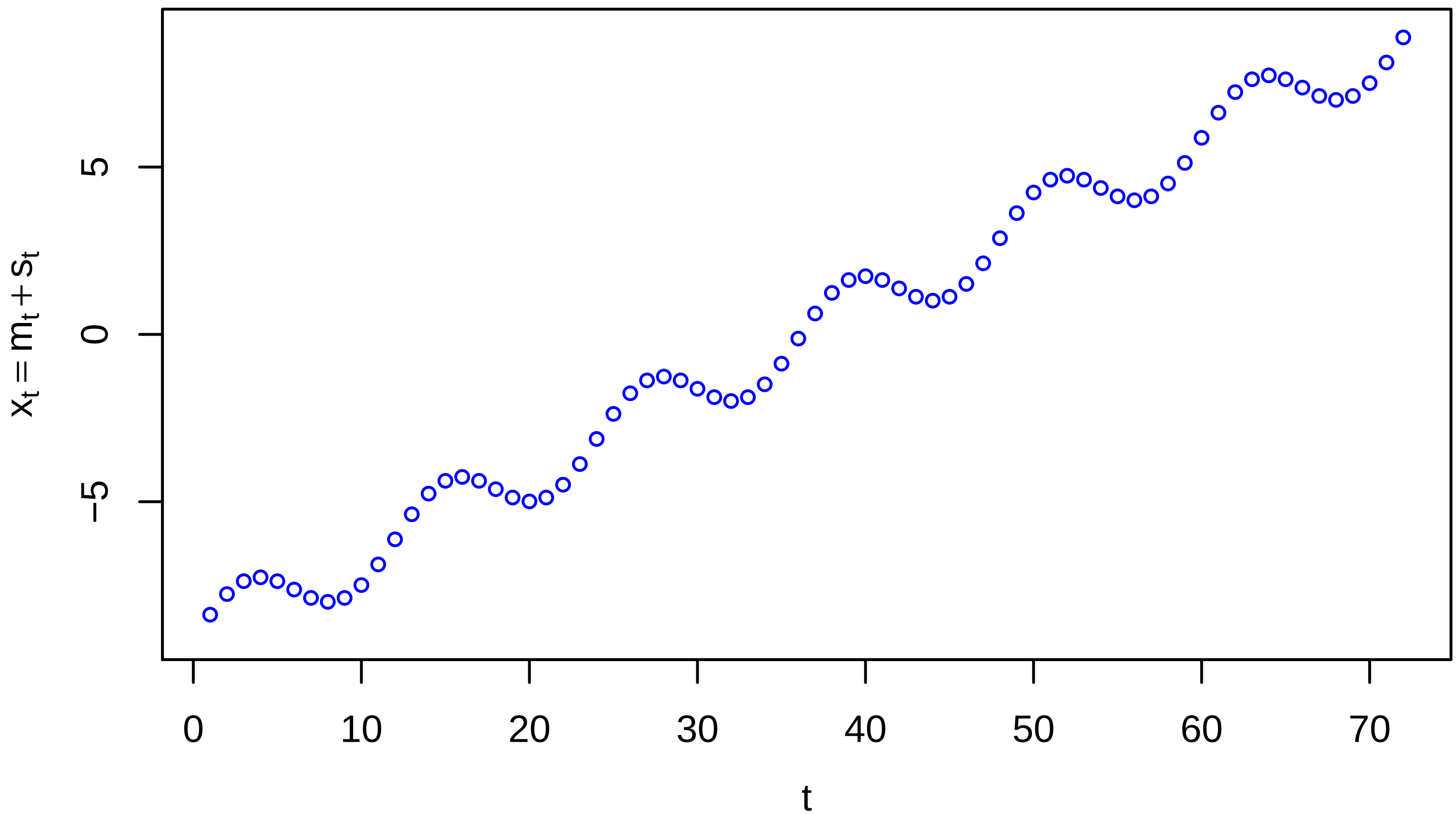
Toy Trend $\{m_t\}$



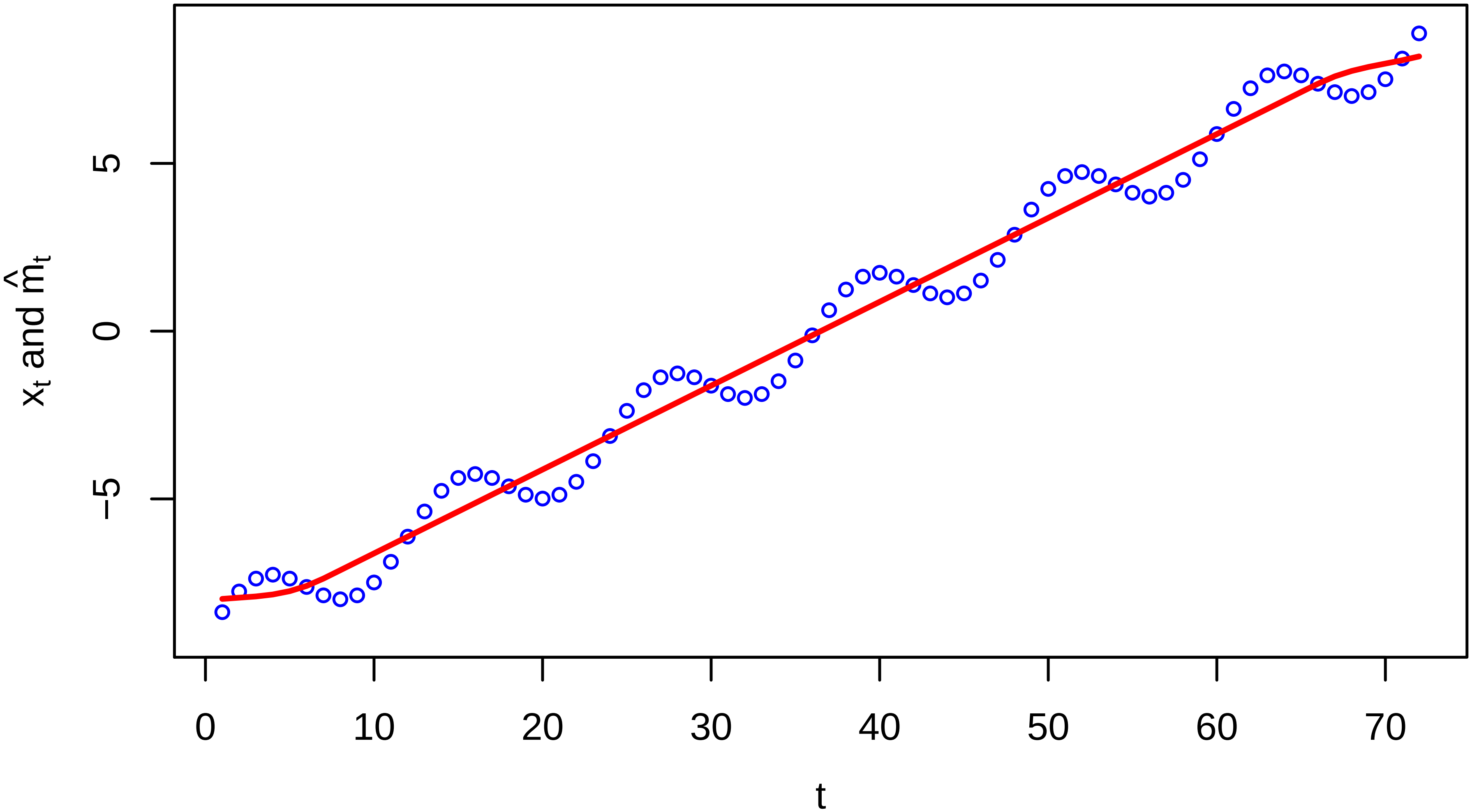
Toy Seasonal Component $\{s_t\}$



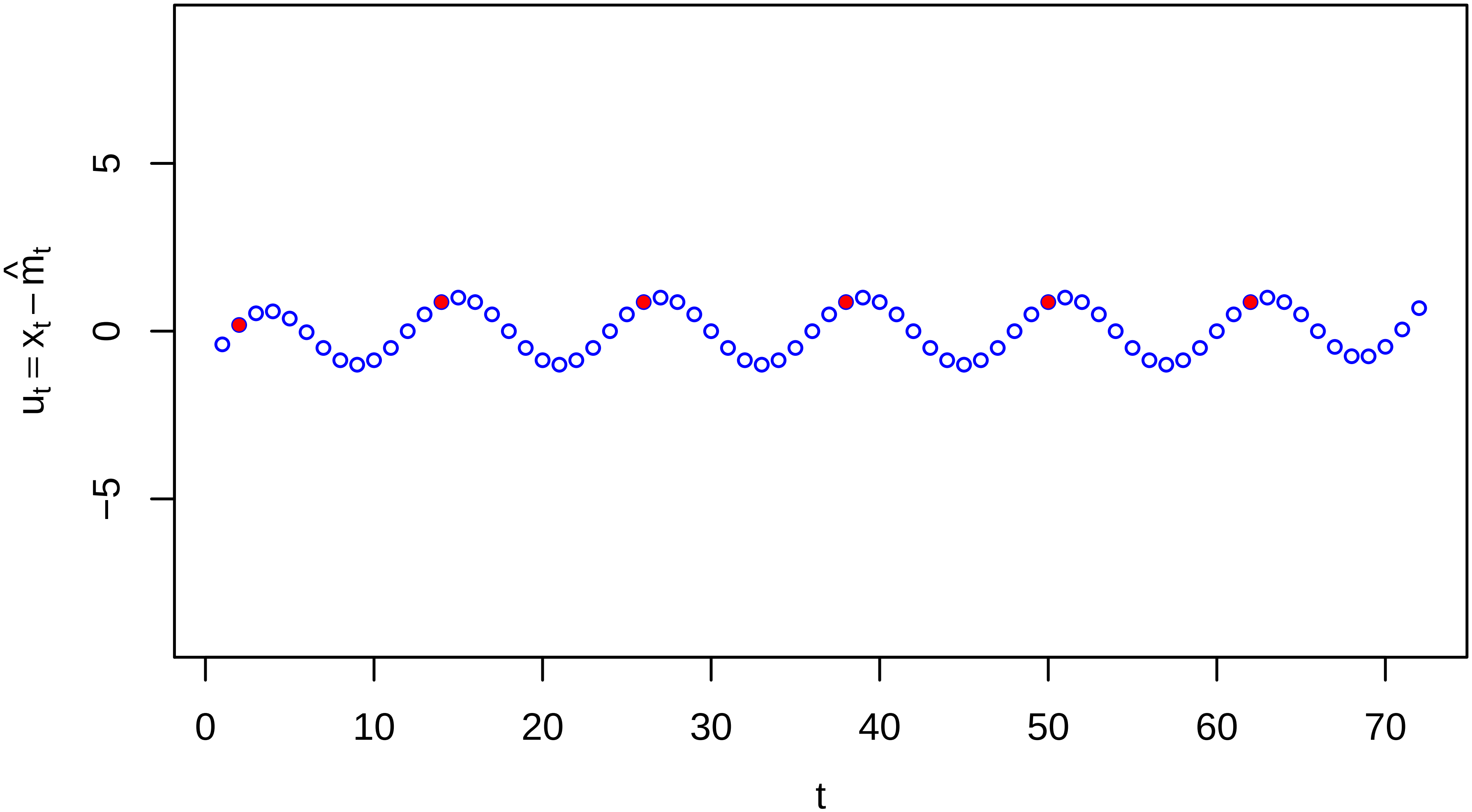
Toy Time Series $\{x_t\}$



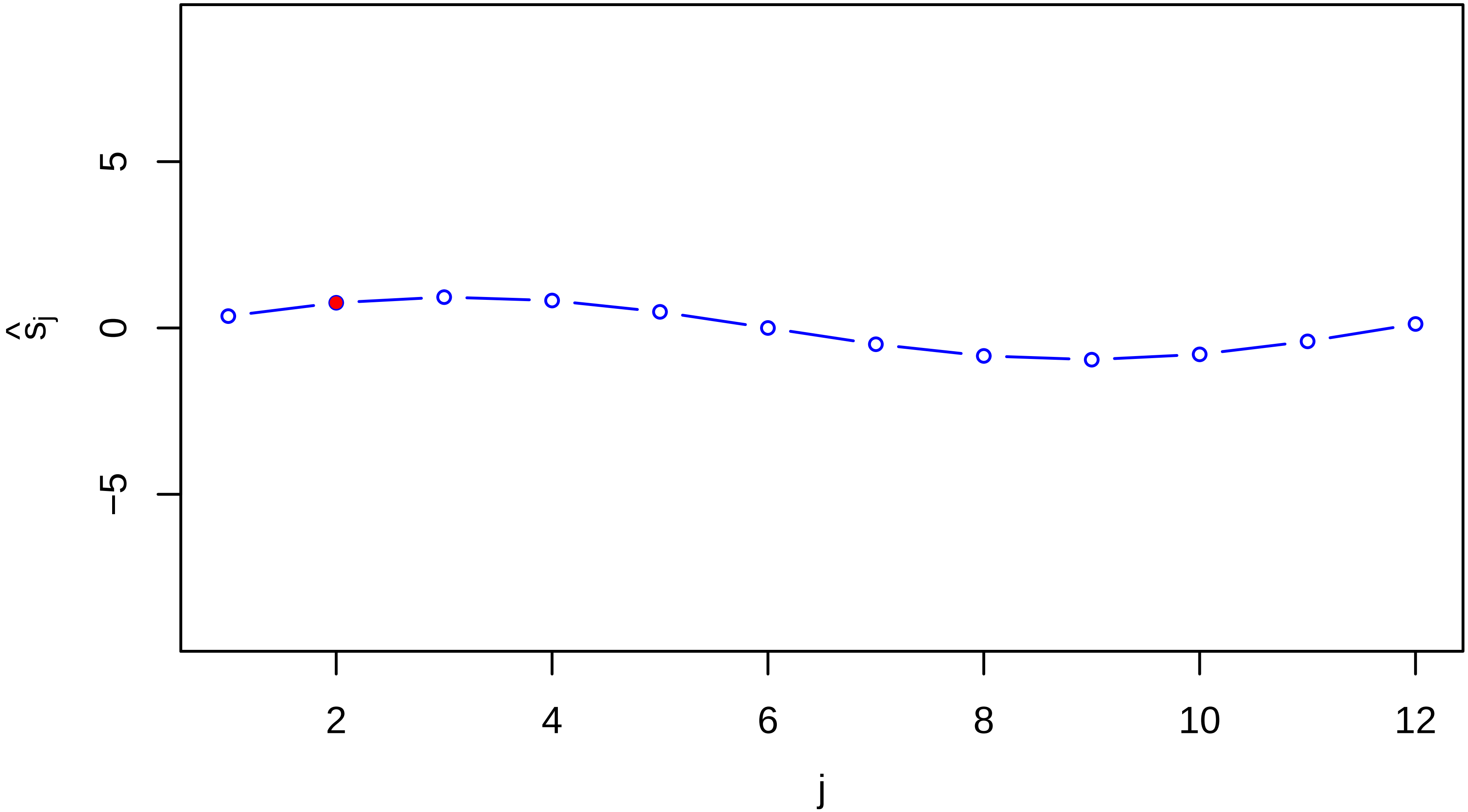
Step 1: Form Preliminary Estimate $\{\hat{m}_t\}$ of Trend



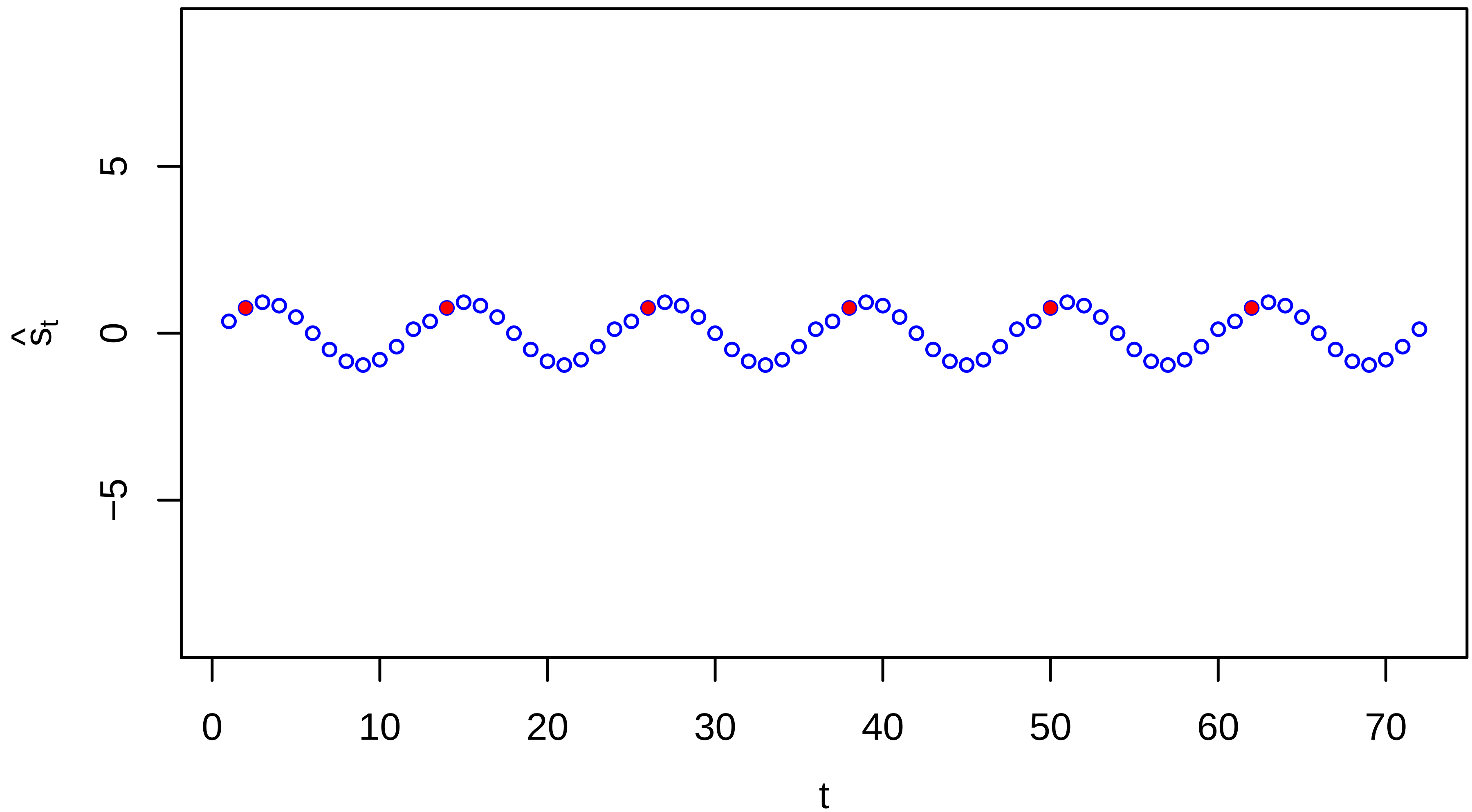
Step 2: Subtract $\{\hat{m}_t\}$ from $\{x_t\}$



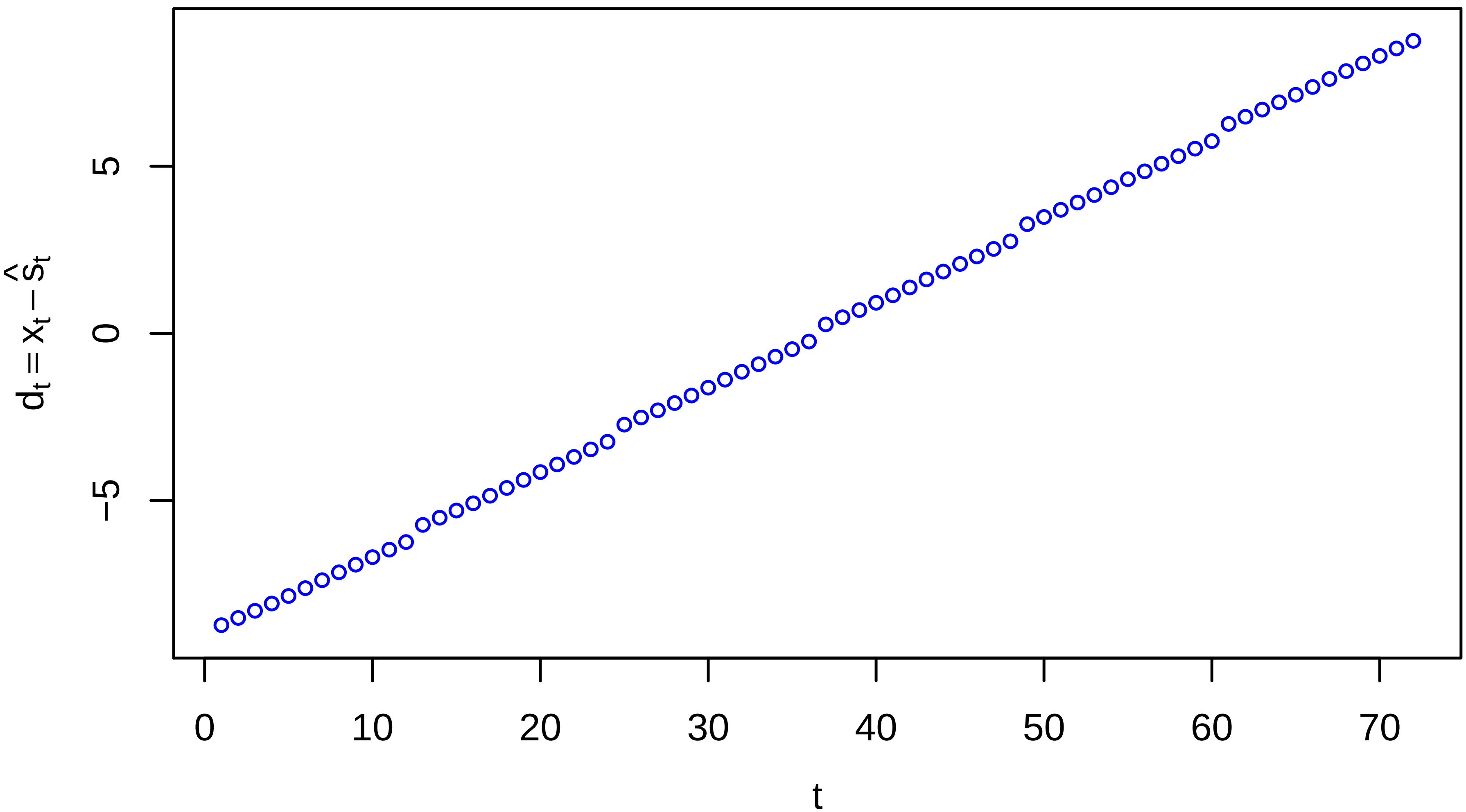
Step 3: Form Estimate $\{\hat{s}_j\}$ of Seasonal Pattern



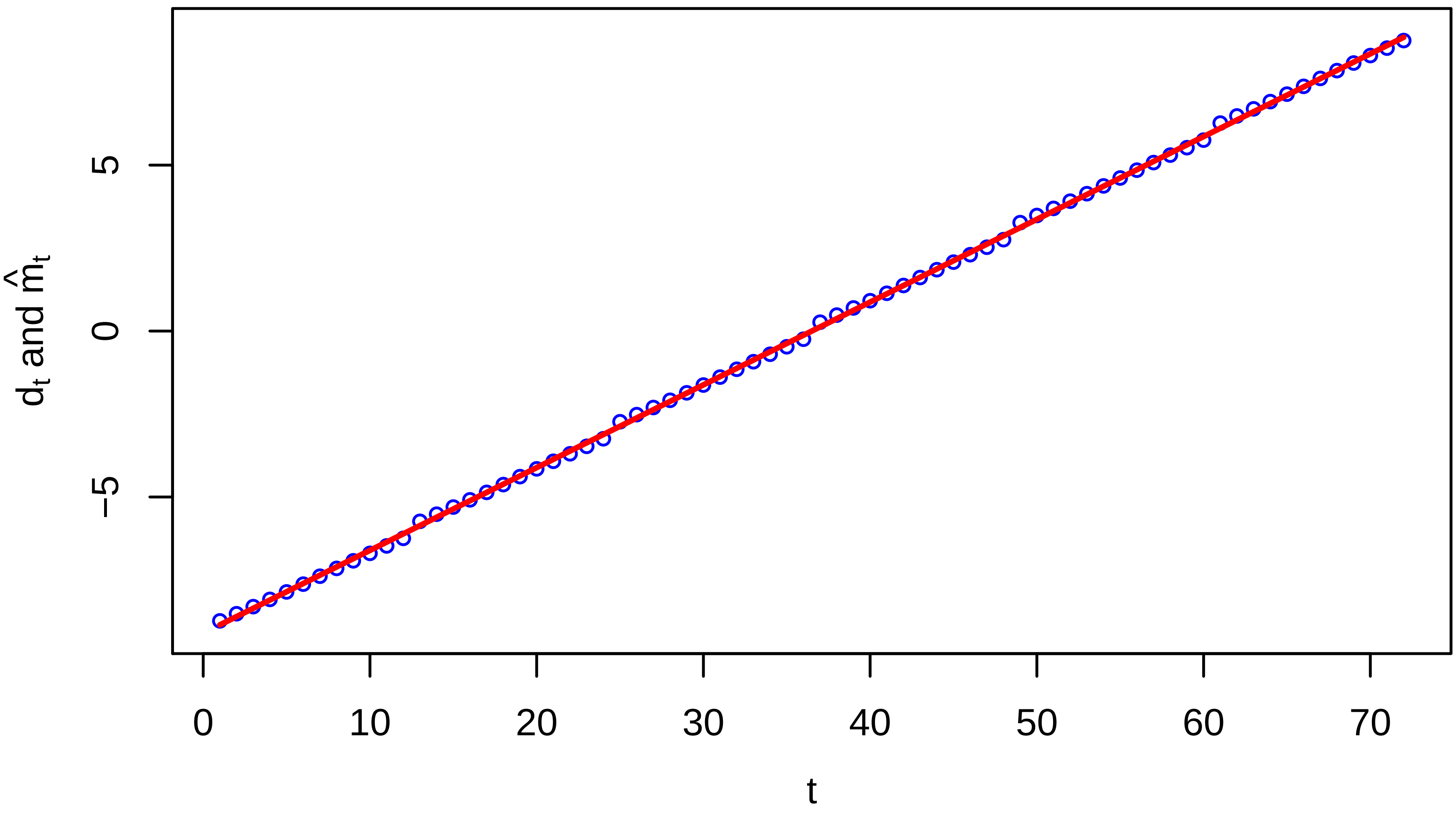
Step 4: Replicate $\{\hat{s}_j\}$ to Form Estimate $\{\hat{s}_t\}$



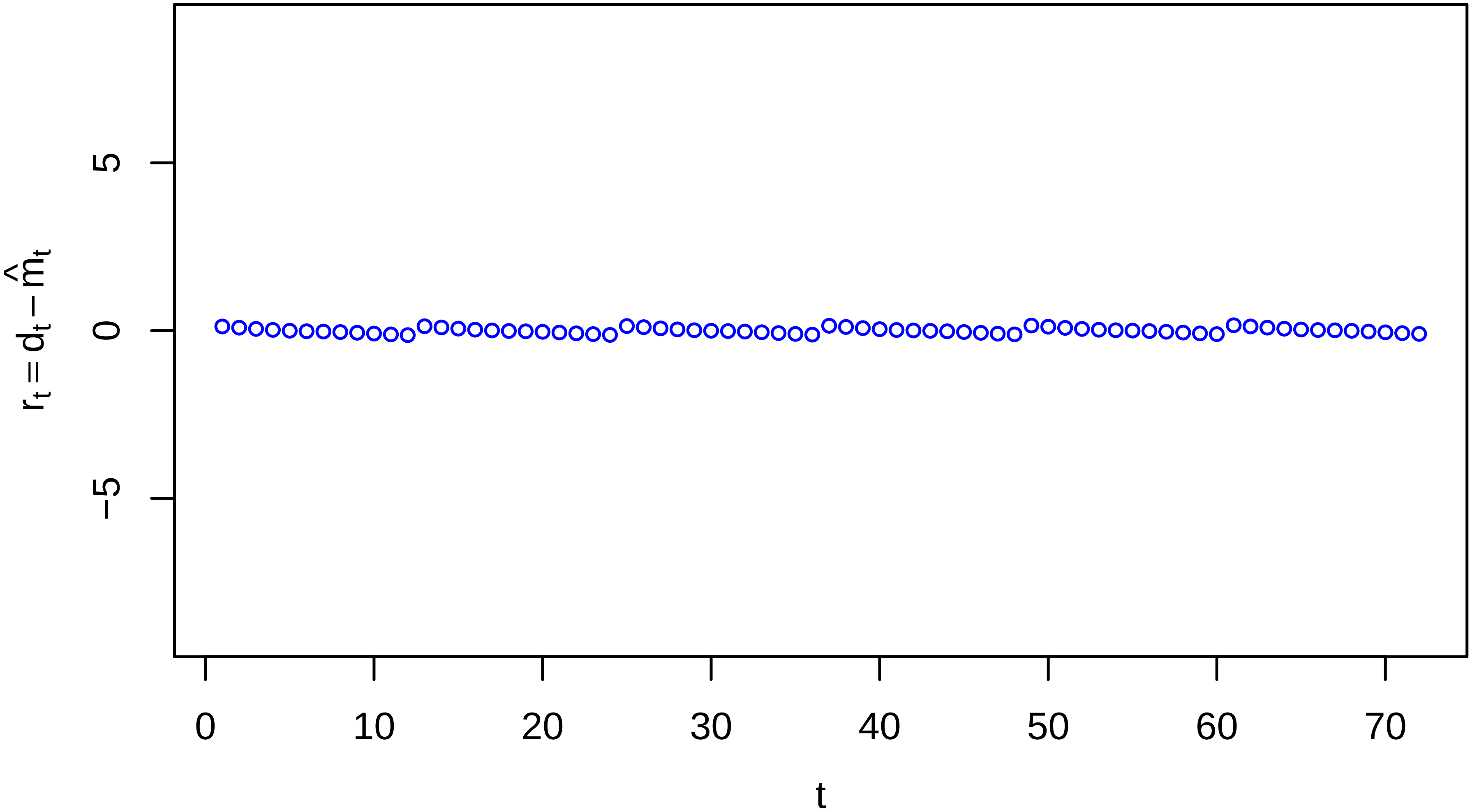
Step 5: Form Deseasonalized Data $d_t = x_t - \hat{s}_t$



Step 6: Fit Line to d_t's to Get Final $\hat{m}_t$'s



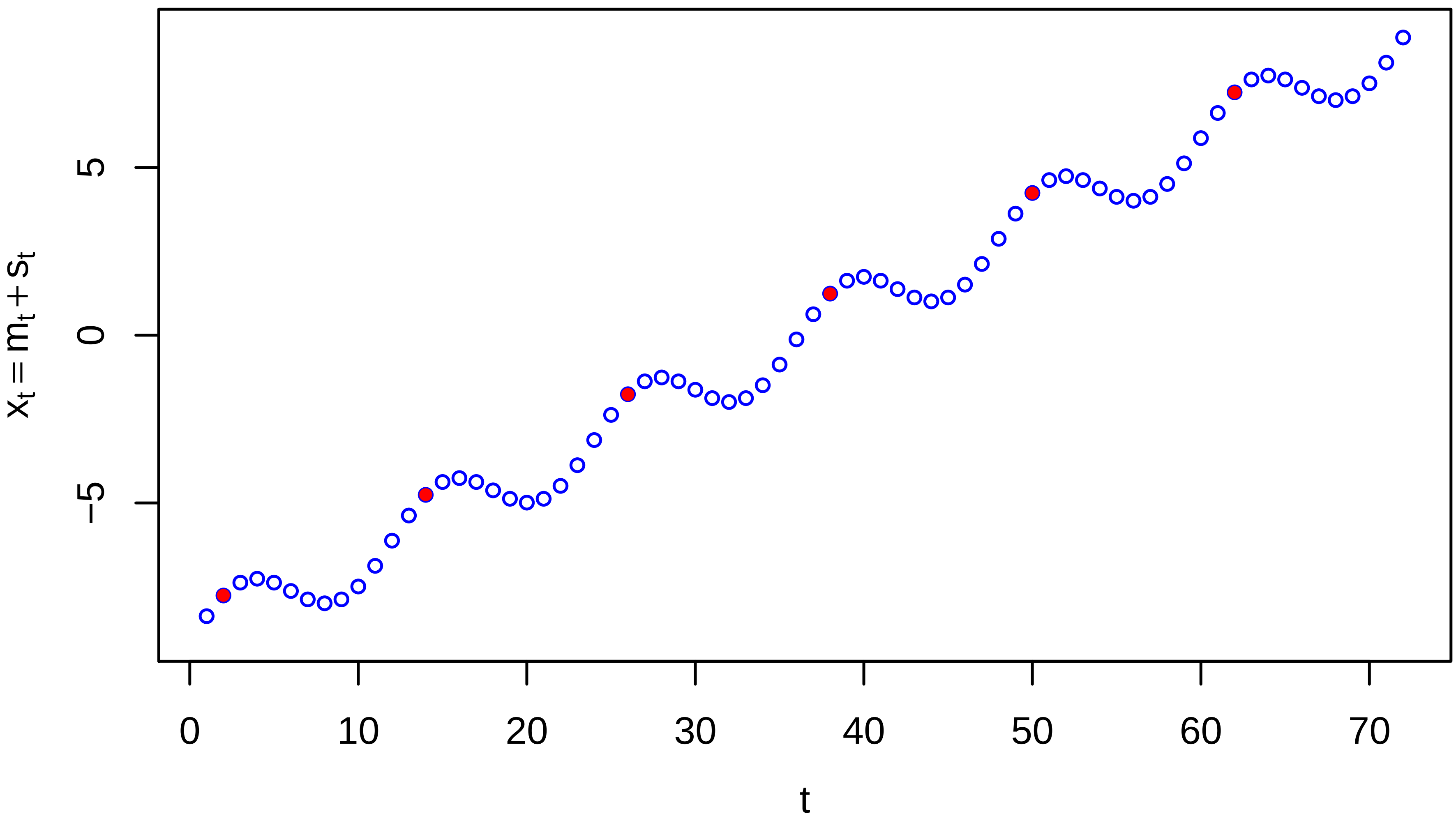
Step 7: Form Residuals from Fitted Line



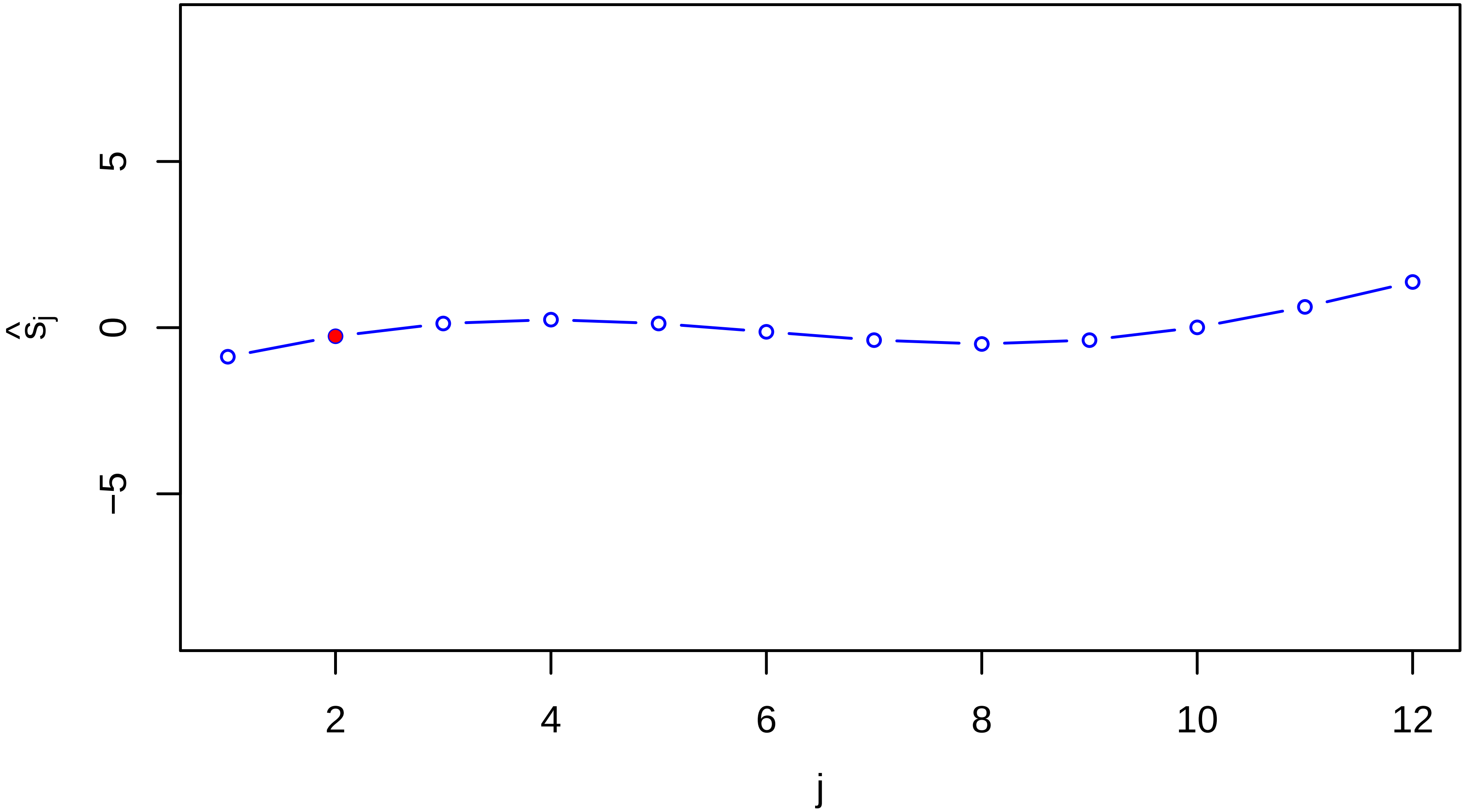
Trend & Seasonal Estimation: IX

- residuals should ideally be zero, but are not quite so due to boundary effects
- now let's see what happens if we eliminate preliminary detrending; i.e., we let $u_t = x_t$ in step 2 and proceed from there

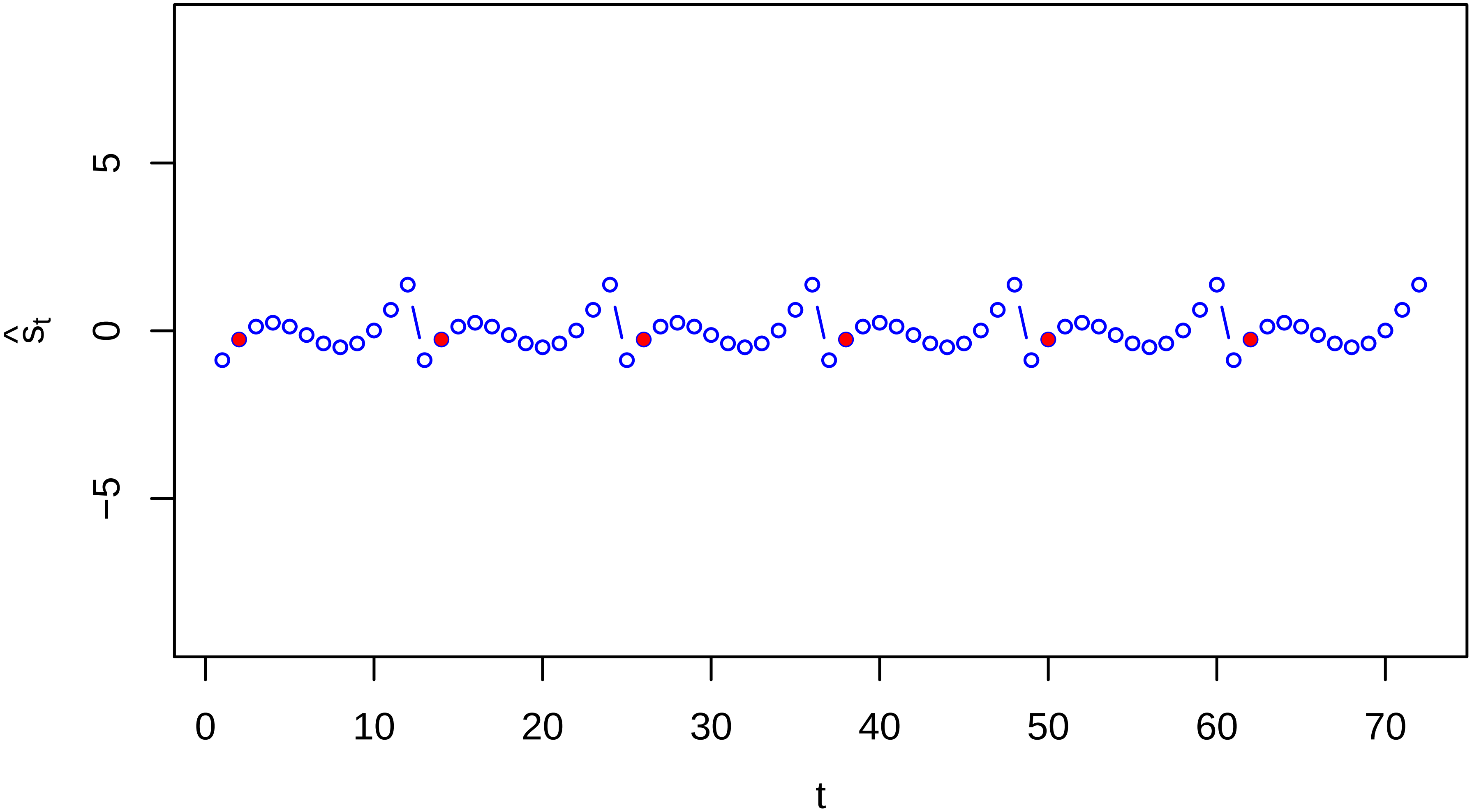
Step 2: Use $\{x_t\}$ in Place of $\{x_t - \hat{m}_t\}$



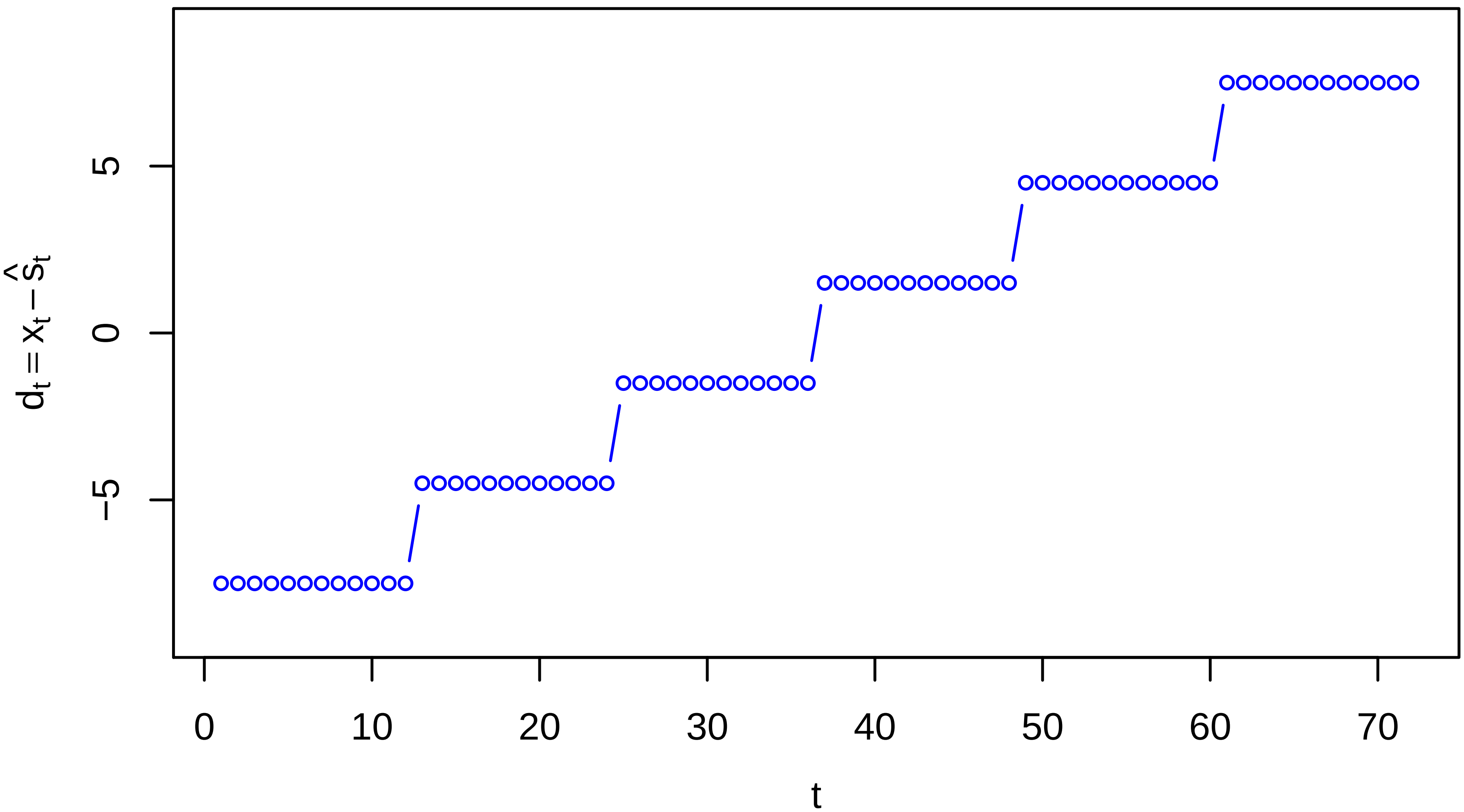
Step 3: Form Estimate $\{\hat{s}_j\}$ of Seasonal Pattern



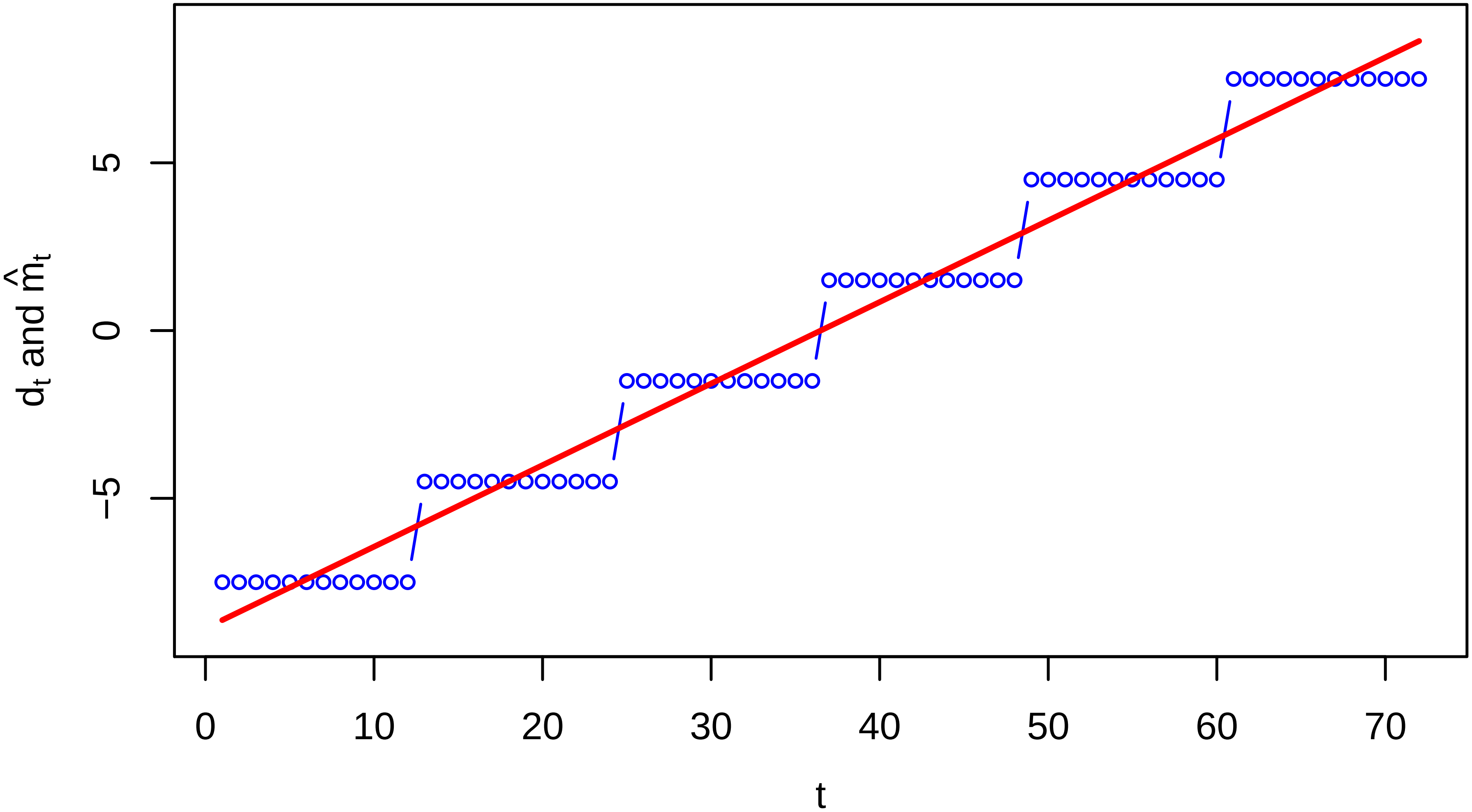
Step 4: Replicate $\{\hat{s}_j\}$ to Form Estimate $\{\hat{s}_t\}$



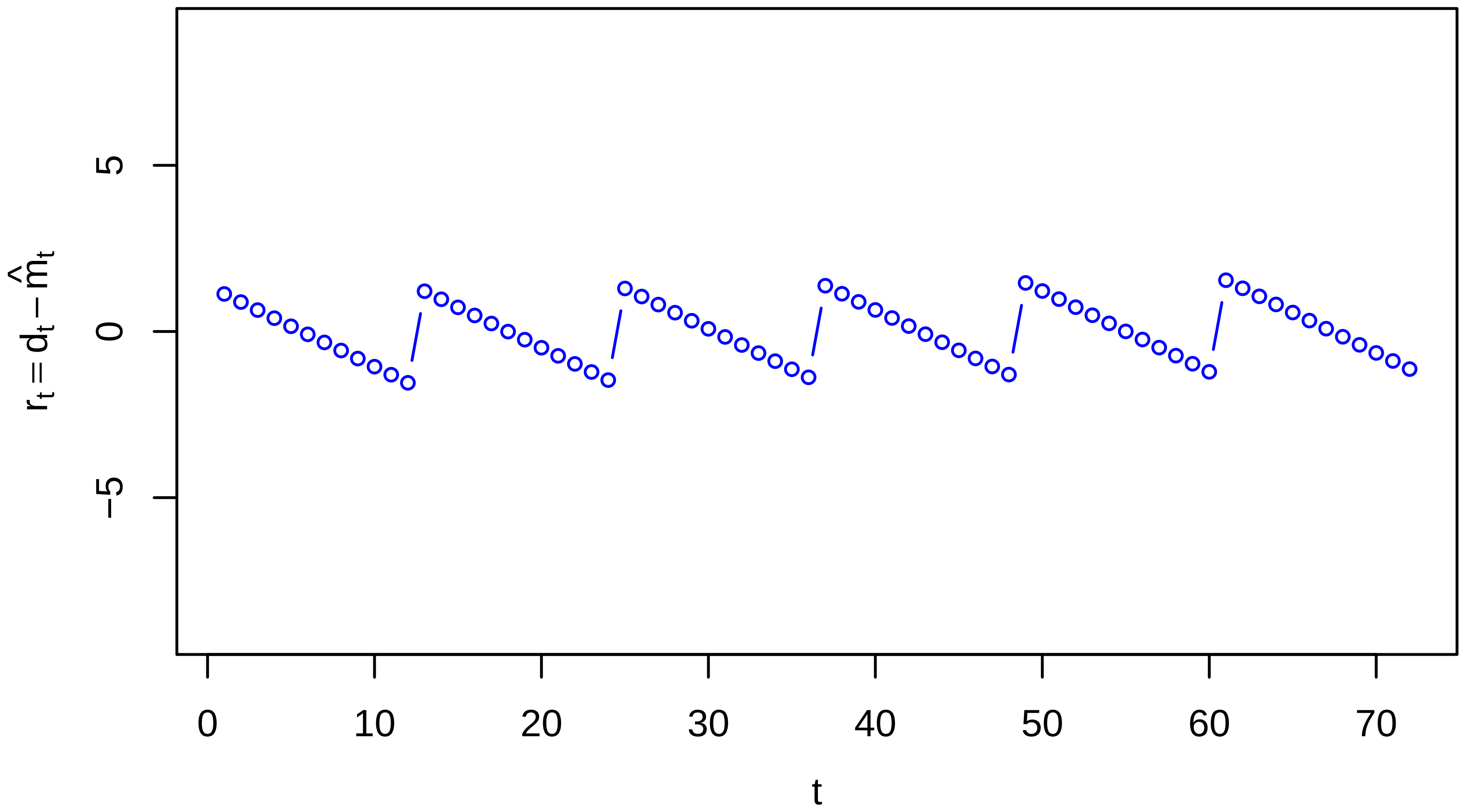
Step 5: Form Deseasonalized Data $d_t = x_t - \hat{s}_t$



Step 6: Fit Line to d_t's to Get Final $\hat{m}_t$'s



Step 7: Form Residuals from Fitted Line



Trend & Seasonal Elimination: I

- let's turn now to the second approach to modeling $\{x_t\}$, which eliminates trend and seasonal components by differencing
- define a lag-d seasonal differencing operator

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d) X_t$$

- application of this operator to model

$$X_t = m_t + s_t + Y_t$$

yields

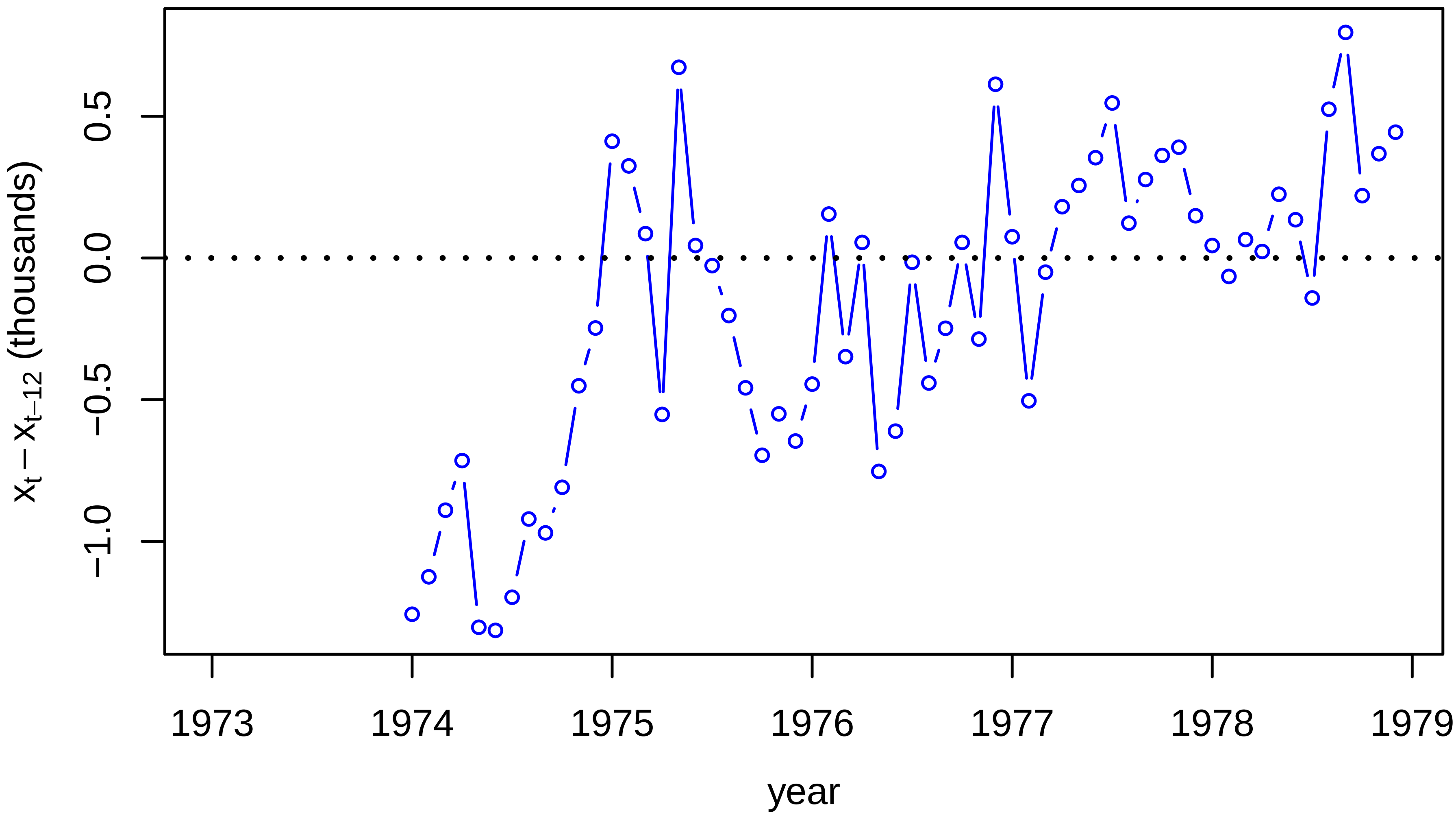
$$\begin{aligned}\nabla_d X_t &= m_t - m_{t-d} + s_t - s_{t-d} + Y_t - Y_{t-d} \\ &= m_t - m_{t-d} + Y_t - Y_{t-d}\end{aligned}$$

because $\{s_t\}$ has period d

Trend & Seasonal Elimination: II

- resulting model $\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$ has a trend component defined by $m_t - m_{t-d}$ and a stochastic component given by $Y_t - Y_{t-d}$
- as before, trend component can be eliminated by applying an appropriate power of operator ∇, say $\nabla^{d'}$
- thus
$$\nabla^{d'} \nabla_d X_t = \nabla^{d'} \nabla_d m_t + \nabla^{d'} \nabla_d Y_t$$
is a model for a series related to $\{x_t\}$ that is free of trend and seasonal components

Accidental Deaths Series After Seasonal Differencing



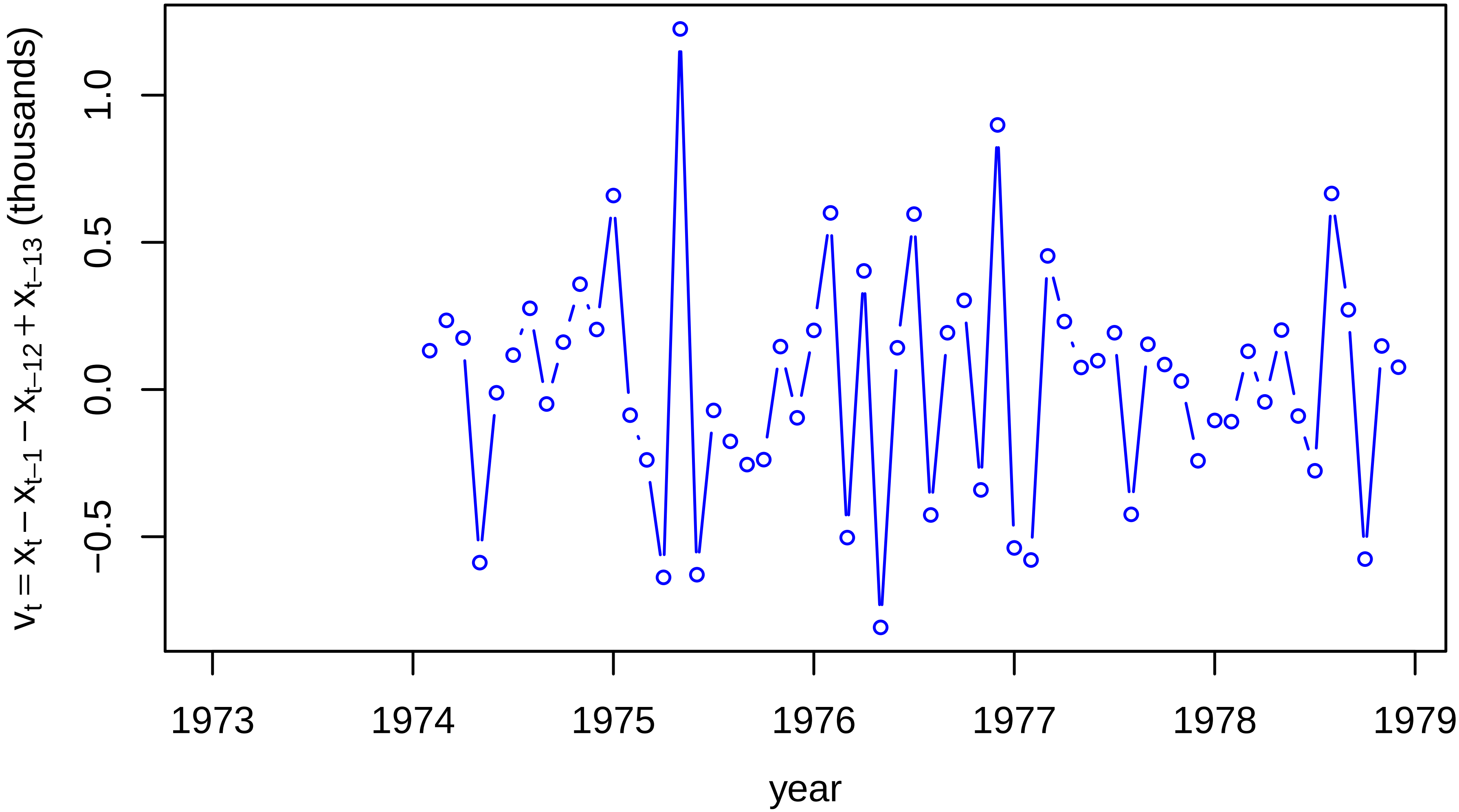
Trend & Seasonal Elimination: III

- Q: how can we interpret upward-looking trend in $x_t - x_{t-12}$?
- model says that

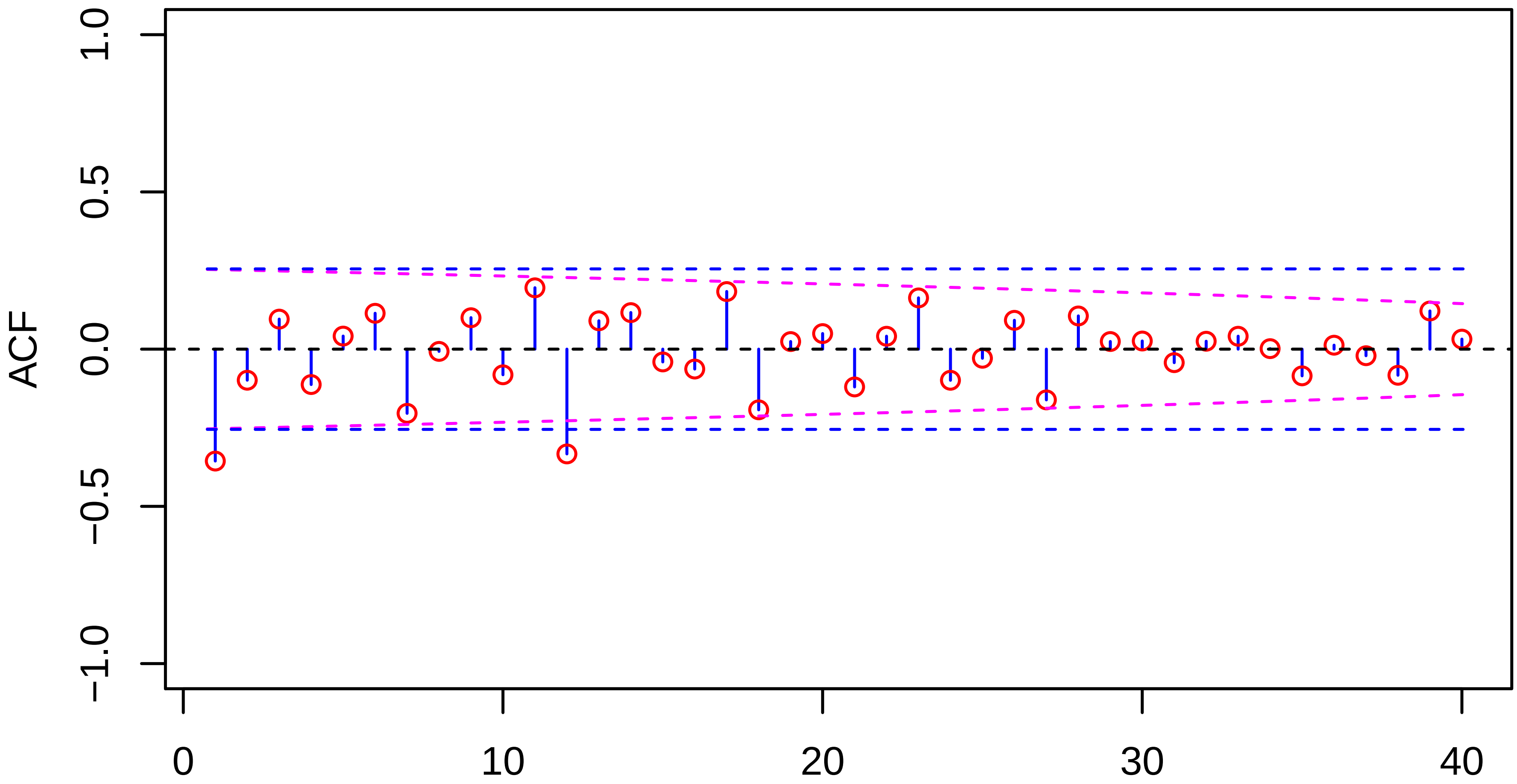
$$\nabla_{12} X_t = X_t - X_{t-12} = m_t - m_{t-12} + Y_t - Y_{t-12},$$

  so trend in $x_t - x_{t-12}$ is $m_t - m_{t-12}$
- if $m_t - m_{t-12} > 0$, trend in x_t has increased over last year
- if $m_t - m_{t-12} < 0$, trend in x_t has decreased over last year
- plot of $x_t - x_{t-12}$ thus suggests that m_t initially decreases (i.e., $x_t - x_{t-12} < 0$) but then increases as we get toward the end of the series (i.e., $x_t - x_{t-12} > 0$)
- this interpretation is consistent with impression we got from preliminary and final estimates of m_t

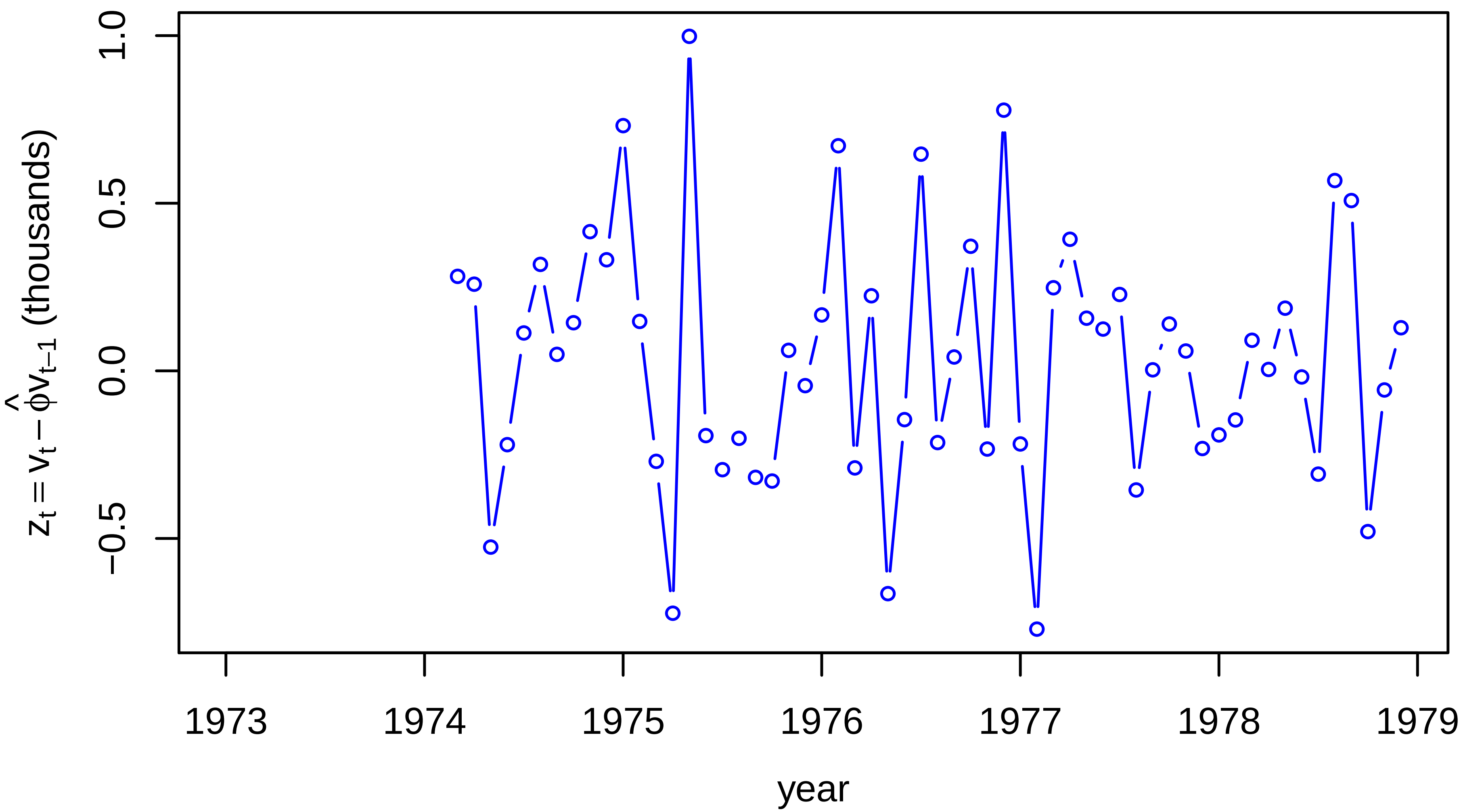
First Difference of Seasonally Differenced $\{x_t\}$ ($\{v_t\}$)



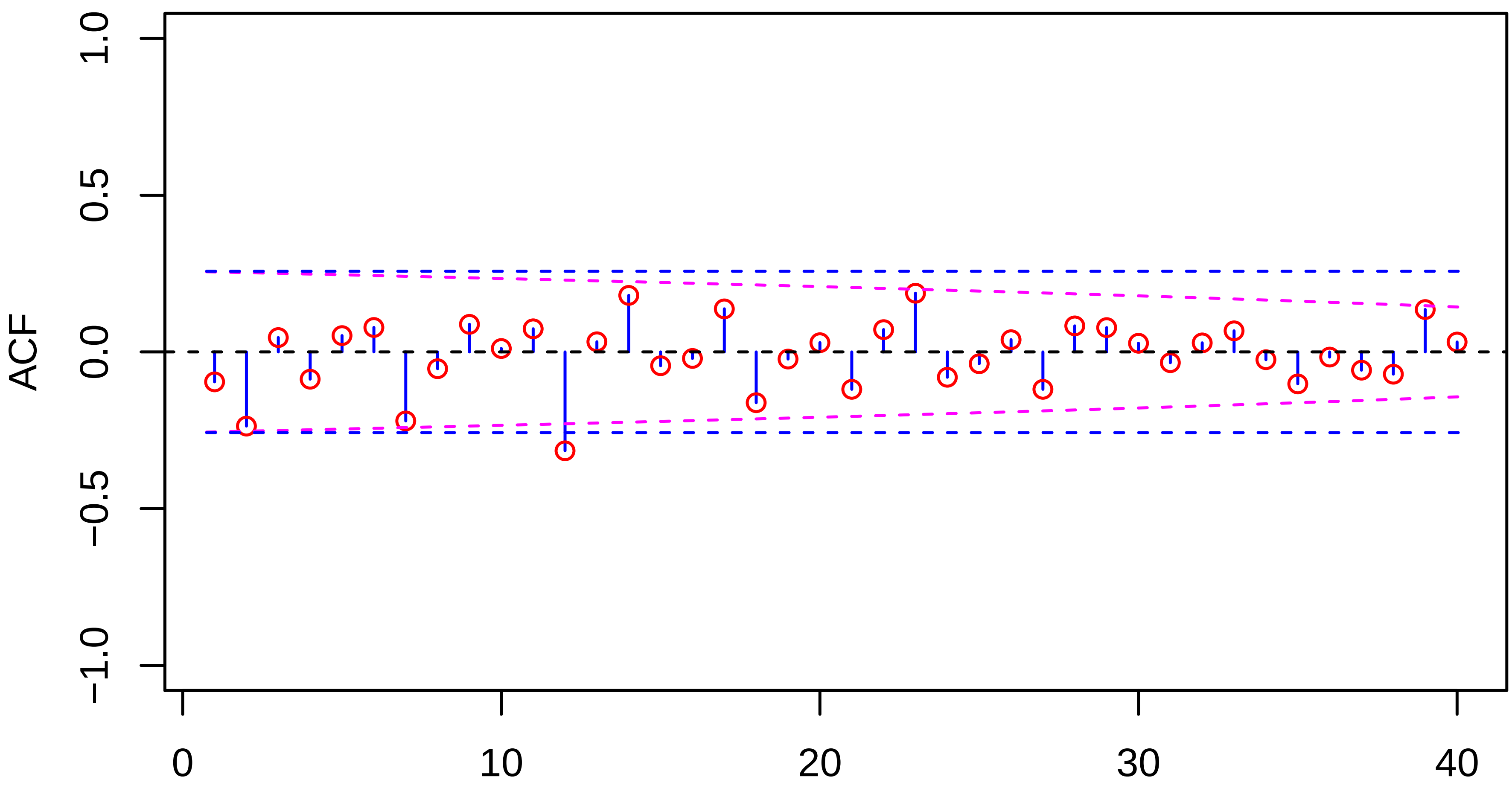
Sample ACF for $\{v_t\}$



Residuals $z_t = v_t - \hat{\phi} v_{t-1}$ from Fitted AR(1) Model



Sample ACF for $\{z_t\}$



Comment on Sample ACF for $\{z_t\}$

- only $\hat{\rho}_z(12)$ outside of $\pm 1.96/\sqrt{n}$ bounds for IID noise
- its p-value is 0.016, which is mildly worrisome, but more so in view of the fact that the lag involved ($h = 12$) is identical to order of seasonal differencing involved ($d = 12$)

Trend/Seasonal Estimation/Elimination – Summary: I

- two simple approaches for using the classical decomposition model
$$X_t = m_t + s_t + Y_t$$
with monthly series $\{x_t\}$ of accidental deaths in USA have yielded two viable models for this time series
- for first model, which estimates $\{m_t\}$ & $\{s_t\}$, needed to
  - estimate seasonal component after preliminary trend removal
  - subtract seasonal component and then reestimate trend
  - fit AR(1) model to what is left over
- for second model, which eliminates $\{m_t\}$ & $\{s_t\}$, needed to
  - apply both seasonal differencing and first differencing
  - fit AR(1) model to what is left over

Trend/Seasonal Estimation/Elimination – Summary: II

- both models are capable of forecasting future values of $\{x_t\}$, but it is not clear at this point which model can be expected to give better forecasts
- model based upon differencing does not give information about $\{m_t\}$ and $\{s_t\}$ directly; alas, these components might be of interest here (and for other time series)
- by contrast, model based on estimating $\{m_t\}$ and $\{s_t\}$ *does* provide this information

References

- P. F. Craigmile, P. Guttorp and D. B. Percival (2004), 'Trend Assessment in a Long Memory Dependence Model Using the Discrete Wavelet Transform,' *Environmetrics*, **15**, pp. 315–35
- R. Hodrick and E. Prescott (1997), 'Postwar U.S. Business Cycles: An Empirical Investigation,' *Money, Credit, and Banking*, **29**, pp. 1–16
- S.–J. Kim, K. Koh, S. Boyd and D. Gorinevsky (2009), 'ℓ_1 Trend Filtering,' *SIAM Review*, **51**, pp. 339–60
- R. L. Smith (1993), 'Long-range Dependence and Global Warming.' In *Statistics for the Environment*, Wiley: New York, pp. 141–61
- R. J. Tibshirani (2014), 'Adaptive Piecewise Polynomial Estimation via Trend Filtering,' *Annals of Statistics*, **42**, pp. 285–323