Climate Time Series Analysis:

Recent Climate Changes

Manfred Mudelsee

Climate Risk Analysis, Germany
Alfred Wegener Institute, Germany
Prologue

Climate: knowledge and experience, space–time scales

Time series analysis: statistical concept

Computer: implementation of concept

Mathematical formulas: no sweat!
Chapters

1. Introduction
2. Persistence Models
3. Bootstrap Confidence Intervals
4. Regression I
5.
6.
7.
8.
9.
1. Introduction

1.1

1.2 Noise and statistical distribution

1.3 Persistence

1.4

1.5 Aim and structure of this course
1. Introduction

Time series sample

\( \{t(i), x(i)\}_{i=1}^{n} \)

- Time series sample
- \( t \), time value
- \( x \), climate-variable value
- \( n \), data size (sample size)
- \( i \), index

**Climate time series**
1. Introduction

Climate equation, continuous time

\[ X(T) = X_{\text{trend}}(T) + X_{\text{out}}(T) + S(T) \cdot X_{\text{noise}}(T), \]  \hspace{1cm} (1.1)

Climate evolves in time, and a stochastic process (a time-dependent random variable representing a climate variable with not exactly known value) and time series (the observed or sampled process) are central to statistical climate analysis. We shall use a wide definition of trend and decompose a stochastic process, \( X \), as follows:

\[ X(T) = X_{\text{trend}}(T) + X_{\text{out}}(T) + S(T) \cdot X_{\text{noise}}(T), \]

where \( T \) is continuous time, \( X_{\text{trend}}(T) \) is the trend process, \( X_{\text{out}}(T) \) is the outlier process, \( S(T) \) is a variability function scaling \( X_{\text{noise}}(T) \), the noise process. The trend is seen to include all systematic or deterministic, long-term processes such as a linear increase, a step change or a seasonal signal. The trend is described by parameters, for example, the rate of an increase. Outliers are events with an extremely large absolute value and are usually rare. The noise process is assumed to be weakly stationary with zero mean and autocorrelation. Giving \( X_{\text{noise}}(T) \) standard deviation unity enables introduction of \( S(T) \) to honour climate's definition as not only the mean but also the variability of the state of the atmosphere and other compartments (Brückner 1890; Hann 1901; Köppen 1923). A version of Eq. (1.1) is written for discrete time, \( T(i) \), as

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \]  \hspace{1cm} (1.2)

using the abbreviation \( X(i) \equiv X(T(i)) \), etc. However, for unevenly spaced \( T(i) \) this is a problematic step because of a possibly non-unique relation between \( X_{\text{noise}}(T) \) and \( X_{\text{noise}}(i) \), see Section 2.1.2.1. The observed, discrete time series from process \( X(i) \) is the set of size \( n \) of paired values \( t(i) \) and \( x(i) \), compactly written as

\[ \{t(i), x(i)\}_{i=1}^{n}. \]

To restate, the aim of this book is to provide methods for obtaining quantitative estimates of parameters of \( X_{\text{trend}}(T) \), \( X_{\text{out}}(T) \), \( S(T) \) and \( X_{\text{noise}}(T) \) using the observed time series data \( \{t(i), x(i)\}_{i=1}^{n} \).

A problem in climate analysis is that the observation process superimposes on the climatic process. \( X_{\text{noise}}(T) \) may show not only climatic
1. Introduction

Climate equation, discrete time

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \quad (1.2) \]

Climate equation, discrete time

\( T(i) \), time points

\( X(i) \equiv X(T(i)) \), etc.
1. Introduction

Stochastic process \( X(i) \)  

“Process level”

Time series \( \{ t(i), x(i) \}_{i=1}^{n} \)  

“Sample level”

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

Univariate: \( T(i), X(i) \)

Bivariate: \( T(i), X(i), Y(i) \)
1. Introduction

1.2 Noise and statistical distribution

The noise, \( X_{\text{noise}}(T) \), has been written in Eq. (1.1) as a zero-mean and unit-standard deviation process, leaving freedom as regards its other second and higher-order statistical moments, which define its distributional shape and also its spectral and persistence properties (next section). The probability density function (PDF), \( f(x) \), defines

\[
\text{prob} (a \leq X_{\text{noise}}(T) \leq a + \delta) = \int_{a}^{a+\delta} f(x) \, dx,
\]

(1.3)

Probability density function (PDF), \( f(x) \)

Gaussian or normal PDF
1. Introduction

1.2 Noise and statistical distribution

Histogram: inference of PDF

Residuals, "detrended version" of the $x(i)$

Figure 1.11. Statistical noise distributions of selected climate time series.

- a: ODP 846 $\delta^{18}O$;
- b: Vostok CO$_2$;
- c: Vostok $\delta^D$;
- d: NGRIP SO$_4$;
- e: NGRIP Ca;
- f: NGRIP dust content;
- g: NGRIP electrical conductivity;
- h: NGRIP Na;
- i: tree-ring $\Delta^{14}C$;
- j: Q5 $\delta^{18}O$;
- k: Lower Mystic Lake varve thickness;
- l: HadCM3 runoff.

The distributions are estimated with histograms. Data and units are given in Figs. 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, and 1.9. In a and e–h, the trend component was estimated (and removed prior to histogram calculation) using a ramp regression model (Figs. 4.6 and 4.7); in b and c using a harmonic filter (Section 5.2.4.3); in d and k using the running median (Figs. 4.16 and 4.17); in i using nonparametric regression (Fig. 4.14); in j using a combination of a ramp model in the early and a sinusoidal in the late part (Fig. 4.18); and in l using the break regression model (Fig. 4.12). Outliers are tentatively marked with open circles (note broken axes in d, k). In c, the modes of the suspected bimodal distribution are marked with arrows. In a, e–h, and j, time-dependent variability, $S(T)$, was estimated using a ramp regression model (Chapter 4); in d and k using the running MAD (Figs. 4.16 and 4.17); and in l using a linear model. Normalizing (dividing by $S(T)$) for those time series was carried out prior to histogram calculation. The other time series assume constant $S(T)$ values are given in Table 1.3.
1. Introduction

1.3 Persistence

\[ E [X_{\text{noise}}(T_1) \cdot X_{\text{noise}}(T_2)] \quad \text{for} \quad T_1 \neq T_2 \]

Autocovariance

\( E, \) expectation operator (averaging)
\( E[X_{\text{noise}}(T)] = 0 \) (zero mean)

Positive autocovariance: persistence or “memory”
1.3 Persistence

Lag-1 scatterplot $x(i-1)$ vs $x(i)$: inference about persistence

Residuals, “detrended version” of the $x(i)$
1. Introduction

1.3 Persistence

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \quad (1.2) \]

\[ \{t(i), x(i)\}_{i=1}^{n}. \]

Residuals:
use sample to determine trend, outliers and variability,
then
subtract trend, remove outliers and divide by variability
(details: Chapters 3 and 4)
1. Introduction

1.5 Aim and structure of this course

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \quad (1.2) \]

Climate equation, stochastic process \( X(i) \)

Time series \( \{ t(i), x(i) \}_{i=1}^{n} \)

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

Univariate: \( T(i), X(i) \)

Bivariate: \( T(i), X(i), Y(i) \)
2. Persistence Models

2.1 First-order autoregressive model

2.2

2.3

2.4 Other models

2.5
2. Persistence Models

2.1 First-order autoregressive model

\[ X_{\text{noise}}(1) = \mathcal{E}_{N(0, 1)}(1), \]
\[ X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0, 1-a^2)}(i), \quad i = 2, \ldots, n. \]  

\( a \) process, even spacing

\(-1 < a < 1\)

Innovation term

\[ \mathcal{E}_{N(\mu, \sigma^2)}(\cdot) \]

Mean = \( \mu = 0 \)

Variance = \( \sigma^2 = 1 \)

Gaussian shape

\[
f(x) = (2\pi)^{-1/2} \sigma^{-1} \exp\left[-(x-\mu)^2/(2\sigma^2)\right]
\]

\[
f(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)
\]
2. Persistence Models

2.1 First-order autoregressive model

AR(1) process, even spacing

\[
X_{\text{noise}}(1) = \mathcal{E}_N(0, 1)(1),
\]
\[
X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_N(0, 1-a^2)(i), \quad i = 2, \ldots, n. \tag{2.1}
\]

Mean: \( E[X_{\text{noise}}(i)] = 0 \) for all \( i \)

Variance: \( \text{VAR}[X_{\text{noise}}(i)] = 1 \) for all \( i \)

Strict stationarity
2. Persistence Models

2.1 First-order autoregressive model

\[ X_{\text{noise}}(1) = \mathcal{E}_{N(0, 1)}(1), \]
\[ X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0, 1-a^2)}(i), \quad i = 2, \ldots, n. \] (2.1)

AR(1) process, even spacing

\[ \rho(h) = \frac{E\left\{ X_{\text{noise}}(i + h) - E\left[ X_{\text{noise}}(i + h) \right] \right\} \cdot \left\{ X_{\text{noise}}(i) - E\left[ X_{\text{noise}}(i) \right] \right\}}{\left\{ \text{VAR}\left[ X_{\text{noise}}(i + h) \right] \cdot \text{VAR}\left[ X_{\text{noise}}(i) \right] \right\}^{1/2}} \]
\[ = E\left[ X_{\text{noise}}(i + h) \cdot X_{\text{noise}}(i) \right], \]

Autocorrelation (lag \( h \))

\[ \rho(h) = a^{\left| h \right|}, \quad h = 0, \pm 1, \pm 2, \ldots. \]
2. Persistence Models

2.1 First-order autoregressive model

\[ X_{\text{noise}}(1) = \mathcal{E}_{N(0, 1)}(1), \]
\[ X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0, 1-a^2)}(i), \quad i = 2, \ldots, n. \] (2.1)

AR(1) process, even spacing

\( a > 0 \): persistence or “memory”

*Figure 2.1.* Realization of an AR(1) process (Eq. 2.1); \( n = 200 \) and \( a = 0.7 \).

*Figure 2.2.* Autocorrelation function of the AR(1) process, \( a > 0 \). In the case of even
2. Persistence Models

2.1 First-order autoregressive model

\[\begin{align*}
X_{\text{noise}}(1) &= \mathcal{E}_{N(0,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0,1-a^2)}(i), \quad i = 2, \ldots, n. 
\end{align*}\] (2.1)

AR(1) process, even spacing

\[\hat{a} = \sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i - 1) \bigg/ \sum_{i=2}^{n} x_{\text{noise}}(i)^2.\] (2.4)

Estimation

Residuals \(x_{\text{noise}}(i)\), “detrended version” of the \(x(i)\)
2. Persistence Models

2.1 First-order autoregressive model

\begin{align*}
X_{\text{noise}}(1) &= \mathcal{E}_{\text{N}(0,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{\text{N}(0,1 - a^2)}(i), \quad i = 2, \ldots, n. \quad (2.1)
\end{align*}

AR(1) process, even spacing

\begin{align*}
\epsilon(i) &= x_{\text{noise}}(i) - \hat{a} \cdot x_{\text{noise}}(i - 1), \quad i = 2, \ldots, n. \quad (2.5)
\end{align*}

White-noise residuals (estimation residuals)

Graphical test of model suitability
2. Persistence Models

2.1 First-order autoregressive model

Artificial time series \((n = 200, a = 0.7)\)
\[ \hat{a} = 0.717 \]

Data from Book, Fig. 2.1
2. Persistence Models

2.1 First-order autoregressive model

Artificial time series \((n = 200, \ a = 0.7)\)

\(\hat{a} = 0.717\)

Theoretically:

\[
E(\hat{a}) \approx (1 - 2/n) a.
\]

Negative bias, underestimation

\[
VAR(\hat{a}) \approx (1 - a^2) / n
\]

Estimation variance

\(VAR[\hat{a}]^{1/2} \approx 0.055\)

Data from Book, Fig. 2.1; Book, Eqs. (2.39), (2.42)
2. Persistence Models

2.1 First-order autoregressive model

\[
\begin{align*}
X_{\text{noise}}(1) &= \mathcal{E}_{N(0,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{N(0,1-a^2)}(i), \quad i = 2, \ldots, n. 
\end{align*}
\] (2.1)

AR(1) process, even spacing

\(a > 0\): persistence,
- fewer independent data points,
- effective data size \(n' < n\),
- larger estimation variance
2. Persistence Models

2.1 First-order autoregressive model

\[
\bar{X} = \sum_{i=1}^{n} X(i)/n,
\]
Mean estimation

\[
n'_\mu = n \left[ 1 + 2 \sum_{i=1}^{n-1} (1 - i/n) \rho(i) \right]^{-1}
\]
Effective data size

![Graph showing effective data size versus data size, n, for various values of a.]
2. Persistence Models

2.4 Other models

Long memory processes

- Fractional Gaussian noise

ARFIMA($p$, $\delta$, $q$)

Long memory

Hyperbolically decreasing acf (for $h \to \infty$)

Mandelbrot (1983), Beran (1994, 1997, 1998), Doukhan et al. (2003), Robinson (2003); Book, Sections 2.4.1, 2.5.2, 2.5.3
2. Persistence Models

2.4 Other models

\[ X_{\text{noise}}(1) = \mathcal{E}_{N(0, 1)}(1), \]
\[ X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0, 1-a^2)}(i), \quad i = 2, \ldots, n. \]  

Gaussian AR(1) process, even spacing

Gaussian innovation term \( \mathcal{E}_{N(\mu, \sigma^2)}(\cdot) \)

Non-Gaussian processes

Non-Gaussian innovation terms
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals
3.2 Bootstrap principle
3.3 Bootstrap resampling
3.4 Bootstrap confidence intervals
3.5
3.6 Bootstrap hypothesis tests
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \]  

(1.2)

Climate equation, stochastic process \(X(i)\)

Time series \(\{t(i), x(i)\}_{i=1}^{n}\)

Time series analysis uses sample to learn about process:
trend parameters, probability of extremes, cycles, etc.
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Estimator

A recipe how to calculate parameter(s) of interest

Estimation

We apply recipe to data and get result.

Estimate

Numerical value of result

Hat notation

Denotes estimator/estimate.
Estimation of AR(1) parameter $a$

$$X_{\text{noise}}(1) = \mathcal{E}_{N(0, 1)}(1),$$
$$X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i - 1) + \mathcal{E}_{N(0, 1-a^2)}(i), \quad i = 2, \ldots, n. \quad (2.1)$$

$$\hat{a} = \frac{\sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i - 1)}{\sum_{i=2}^{n} x_{\text{noise}}(i)^2}. \quad (2.4)$$
3. Bootstrap Confidence Intervals
3.1 Error bars and confidence intervals

True parameter  Sample  Estimator, estimate
Time series

\( \theta \)  \( \{ t(i), x(i) \}_{i=1}^{n} \)  \( \hat{\theta} \)

("Theta")  ("Theta hat")

\( n < \infty \),
Observation errors
(measurement, proxy, calibration, climate model, etc.)

\( \hat{\theta} \neq \theta \)
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Estimate does not equal true (but unknown) value, it deviates.

However, we can say how large this deviation typically is: error bars, confidence intervals, etc.

We need statistical language, we make a statistical inference.
Assumptions implicitly made so far include:

1. Truth exists (logic, realism)
2. Time arrow
3. Ergodicity
4. Foundation of probability (axiomatic approach)

Recommendation:
Be cautious with “blog science” or “post-normal” stuff.
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

PDF of an estimator

\[
\text{se}_{\hat{\theta}} = \left[ \text{VAR} \left( \hat{\theta} \right) \right]^{1/2}.
\]

Standard error

\[
\text{bias}_{\hat{\theta}} = E \left( \hat{\theta} \right) - \theta.
\]

Bias

\[
\text{RMSE}_{\hat{\theta}} = \left\{ E \left[ \left( \hat{\theta} - \theta \right)^2 \right] \right\}^{1/2}
= \left( \text{se}_{\hat{\theta}}^2 + \text{bias}_{\hat{\theta}}^2 \right)^{1/2}.
\]

Root mean squared error

Confidence interval (e.g., equi-tailed)

\[
\text{CI}_{\hat{\theta}, 1-2\alpha} = \left[ \hat{\theta}_l; \hat{\theta}_u \right],
\]
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_{N(\mu, \sigma^2)}(i), \quad i = 1, \ldots, n, \]

\[
\hat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n. 
\]

What is the expectation?
Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_{N(\mu, \sigma^2)}(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X(i). \]

What is the expectation?

\[ (\mu + \mu + \mu + \ldots + \mu)/n = \mu \times n/n = \mu \quad \text{(zero bias)} \]
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise

\[ X(i) = E_{N(\mu, \sigma^2)}(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n. \]

What is the variance?
Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_N(\mu, \sigma^2)(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X(i). \]

What is the variance?

\[ (\sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2)/n^2 = \sigma^2 \times n/n^2 = \sigma^2/n \quad \text{(standard error } \sigma n^{-1/2}) \]
Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_N(\mu, \sigma^2)(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n. \]

What is the distribution of the estimator?
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_{N(\mu, \sigma^2)}(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X(i). \]

What is the distribution of the estimator?

Student’s t-distribution (we can calculate CIs)

Exact CIs have coverages, equal to the nominal value 1. Constructions of exact CIs require knowledge of the distribution of the estimator, which is ideally small. As a second CI property besides coverage, we consider interval length, bias. The properties of interest for CIs are the coverages, RMSE, \( \gamma \). An example is the autocorrelation function, which is called a Gaussian purely random process or Gaussian white noise. Following the usual convention, we denote also the distribution of the estimator (Eq.3.19). The properties of interest for CIs are the coverages, RMSE, \( \gamma \).
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Table 3.1. Monte Carlo experiment, mean estimation of a Gaussian purely random process. \( n_{\text{sim}} = 4,750,000 \) random samples of \( \{X(i)\}_{i=1}^{n} \) were generated after Eq. (3.9) with \( \mu = 1.0, \sigma = 2.0 \) and various \( n \) values. An exact confidence interval \( \text{CI}_{\bar{x},1-2\alpha} \) was constructed for each simulation after Eq. (3.18) with \( \alpha = 0.025 \). Average CI length, empirical RMSE \( \bar{x} \) and empirical coverage were determined subsequently. The entries are rounded.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}_{\bar{x}}^a )</th>
<th>( \text{Nominal}^b )</th>
<th>( \langle \text{CI length} \rangle^c )</th>
<th>( \text{Nominal}^d )</th>
<th>( \gamma_{\bar{x}}^e )</th>
<th>( \text{Nominal} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.6327</td>
<td>0.6325</td>
<td>2.7832</td>
<td>2.7832</td>
<td>0.9499</td>
<td>0.9500</td>
</tr>
<tr>
<td>20</td>
<td>0.4474</td>
<td>0.4472</td>
<td>1.8476</td>
<td>1.8476</td>
<td>0.9498</td>
<td>0.9500</td>
</tr>
<tr>
<td>50</td>
<td>0.2828</td>
<td>0.2828</td>
<td>1.1310</td>
<td>1.1310</td>
<td>0.9501</td>
<td>0.9500</td>
</tr>
<tr>
<td>100</td>
<td>0.2000</td>
<td>0.2000</td>
<td>0.7916</td>
<td>0.7917</td>
<td>0.9499</td>
<td>0.9500</td>
</tr>
<tr>
<td>200</td>
<td>0.1415</td>
<td>0.1414</td>
<td>0.5570</td>
<td>0.5571</td>
<td>0.9499</td>
<td>0.9500</td>
</tr>
<tr>
<td>500</td>
<td>0.0894</td>
<td>0.0894</td>
<td>0.3513</td>
<td>0.3513</td>
<td>0.9500</td>
<td>0.9500</td>
</tr>
<tr>
<td>1000</td>
<td>0.0633</td>
<td>0.0632</td>
<td>0.2482</td>
<td>0.2482</td>
<td>0.9499</td>
<td>0.9500</td>
</tr>
</tbody>
</table>

\( a \) Empirical RMSE \( \bar{x} \), given by \( \left[ \sum_{i=1}^{n_{\text{sim}}} (\bar{x} - \mu)^2 / n_{\text{sim}} \right]^{1/2} \).

\( b \) \( \sigma \cdot n^{-1/2} \).

\( c \) Average value over \( n_{\text{sim}} \) simulations.

\( d \) \( 2 \cdot t_{n-1}(1-\alpha) \cdot \sigma \cdot c \cdot n^{-1/2} \), where \( c \) is given by Eq. (3.24).

\( e \) Empirical coverage, given by the number of simulations where \( \text{CI}_{\bar{x},1-2\alpha} \) contains \( \mu \), divided by \( n_{\text{sim}} \). Standard error of \( \gamma_{\bar{x}} \) is (Efron and Tibshirani 1993) nominally \( [2\alpha(1-2\alpha)/n_{\text{sim}}]^{1/2} = 0.0001 \).
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_{N(\mu, \sigma^2)}(i), \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n. \]

Simple data generating process,
simple estimation problem:

we can write down PDF of estimator (i.e., we have everything),
CIs are exact.
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Standard deviation estimation of Gaussian white noise

\[ X(i) = \mathcal{E}_N(\mu, \sigma^2)(i), \quad i = 1, \ldots, n, \]

\[ \widehat{\sigma} = S_{n-1} = \left\{ \sum_{i=1}^{n} \left[ X(i) - \bar{X} \right]^2 / (n - 1) \right\}^{1/2} \]

Simple data generating process, simple estimation problem:
we can write down PDF of estimator (i.e., we have everything), CIs are exact.
3. Bootstrap Confidence Intervals
3.1 Error bars and confidence intervals

Mean estimation of lognormal white noise

\[ X(i) = \exp \left[ \mathcal{E}_{N(\mu, \sigma^2)}(i) \right], \quad i = 1, \ldots, n, \]

\[ \hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X(i)}{n}. \]

Distributional assumption violated:

data generating process lognormal, not Gaussian.

What happens to CI (from Student’s \( t \)-distribution)?
Table 3.3. Monte Carlo experiment, mean and median estimation of a lognormal purely random process. $n_{\text{sim}} = 4,750,000$ random samples of \( \{X(i)\}_{i=1}^{n} \) were generated after \( X(i) = \exp[\mathcal{E}_{N(\mu, \sigma^2)}(i)] \), \( i = 1, \ldots, n \), with \( \mu = 1.0, \sigma = 1.0 \) and various \( n \) values. The density function is skewed (Fig. 3.2). Analysed as estimators of the centre of location of the distribution were the sample mean (Eq. 3.16) and the sample median, \( \hat{m} \) (see background material, Section 3.8). CI $\bar{x}, 1 - 2\alpha$ was constructed after Eq. (3.18) with $\alpha = 0.025$.

<table>
<thead>
<tr>
<th>( n )</th>
<th>$\gamma_{\bar{x}}^a$</th>
<th>Nominal</th>
<th>$C^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.8392</td>
<td>0.9500</td>
<td>$-0.1108$</td>
</tr>
<tr>
<td>20</td>
<td>0.8670</td>
<td>0.9500</td>
<td>$-0.0830$</td>
</tr>
<tr>
<td>50</td>
<td>0.8991</td>
<td>0.9500</td>
<td>$-0.0509$</td>
</tr>
<tr>
<td>100</td>
<td>0.9170</td>
<td>0.9500</td>
<td>$-0.0330$</td>
</tr>
<tr>
<td>200</td>
<td>0.9296</td>
<td>0.9500</td>
<td>$-0.0204$</td>
</tr>
<tr>
<td>500</td>
<td>0.9399</td>
<td>0.9500</td>
<td>$-0.0101$</td>
</tr>
<tr>
<td>1000</td>
<td>0.9442</td>
<td>0.9500</td>
<td>$-0.0058$</td>
</tr>
</tbody>
</table>

$\gamma_{\bar{x}}^a$ is nominally 0.0001. $C^b$ Empirical coverage error of CI $\bar{x}, 1 - 2\alpha$, given by $\gamma_{\bar{x}}$ minus nominal value.
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Real world

Violated assumptions (e.g., distributional shape)
CIs not exact, but only approximate
CIs have coverage error, $C$

$C$ decreases with $n$

- $C \sim 1/n^{1/2}$ first-order accurate CI
- $C \sim 1/n$ second-order accurate CI
3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

Real world

- Complex processes (distributional shape, persistence, spacing)
- Complex estimation problems
- PDF of estimator cannot be derived analytically
- Approximate CIs

Bootstrap

- Real world
- Method to construct (approximate) CIs
3. Bootstrap Confidence Intervals

3.2 Bootstrap principle

\[
\{t^*(i), x^*(i)\}_{i=1}^{n}
\]

\[
\{t^*(i), x^*(i)\}_{i=1}^{n}
\]

\[
\{t^*(i), x^*(i)\}_{i=1}^{n}
\]

\[
\{t^*(i), x^*(i)\}_{i=1}^{n}
\]

\[
\hat{\theta}^*B
\]

\[
\hat{\theta}^*2
\]

\[
\hat{\theta}^*1
\]

\[
\hat{\theta}
\]

\[
\text{CI}_{\hat{\theta},1-2\alpha}
\]

Confidence interval

Replications

Resamples

Sample, estimate

Book, Fig. 3.3
3. Bootstrap Confidence Intervals

3.2 Bootstrap principle

Bootstrap
Resampling
   Moving block bootstrap (MBB)
   Autoregressive bootstrap (ARB)
Other
Confidence interval construction
   Normal
   Student’s t
   Percentile
   BCa
3. Bootstrap Confidence Intervals

3.3 Bootstrap resampling

Moving block bootstrap (MBB)

**Algorithm 3.1.** Moving block bootstrap algorithm (MBB). Note: An equation like \{t^*_{(i)}\}_{i=1}^n = \{t_{(i)}\}_{i=1}^n is used to denote \(t^*_{(i)} = t_{(i)}, i = 1, \ldots, n\).
# 3. Bootstrap Confidence Intervals

## 3.3 Bootstrap resampling

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Data</th>
<th>{t(i), x(i)}_{i=1}^{n}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2</td>
<td>Resampled times</td>
<td>{t^*(i)}<em>{i=1}^{n} = {t(i)}</em>{i=1}^{n}</td>
</tr>
<tr>
<td>Step 3</td>
<td>Residuals (Eq. 3.29)</td>
<td>( r(i) = [x(i) - \hat{x}<em>{\text{trend}}(i) - \hat{x}</em>{\text{out}}(i)] / \hat{S}(i) )</td>
</tr>
<tr>
<td>Step 4</td>
<td>Apply MBB (Algorithm 3.1)</td>
<td>{r(i)}_{i=1}^{n}</td>
</tr>
<tr>
<td>Step 5</td>
<td>Resampled residuals</td>
<td>{r^*(i)}_{i=1}^{n}</td>
</tr>
<tr>
<td>Step 6</td>
<td>Use resampled residuals to produce resamples</td>
<td>( x^<em>(i) = \hat{x}<em>{\text{trend}}(i) + \hat{x}</em>{\text{out}}(i) + \hat{S}(i) \cdot r^</em>(i) )</td>
</tr>
</tbody>
</table>

**Algorithm 3.3.** MBB for realistic climate processes, which comprise trend, outlier and variability components.
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

$\{t(i), x(i)\}_{i=1}^{n}$

$\{t^{*1}(i), x^{*1}(i)\}_{i=1}^{n}$

$\{t^{*2}(i), x^{*2}(i)\}_{i=1}^{n}$

$\{t^{*B}(i), x^{*B}(i)\}_{i=1}^{n}$

$\hat{\theta}$

$\hat{\theta}^{*1}$

$\hat{\theta}^{*2}$

$\hat{\theta}^{*B}$

$\text{CI}_{\hat{\theta}, 1-2\alpha}$

Confidence interval

Replications

Resamples

Sample, estimate

Book, Fig. 3.3
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

Bootstrap standard error

\[ \text{se}_{\hat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[ \hat{\theta}^{*b} - \langle \hat{\theta}^{*b} \rangle \right]^2 / (B - 1) \right\}^{1/2}, \]  

(3.30)
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

Normal confidence interval

\[
\text{CI}_{\hat{\theta}, 1-2\alpha} = \left[ \hat{\theta} + z(\alpha) \cdot \hat{s}e_{\hat{\theta}^*}; \hat{\theta} - z(\alpha) \cdot \hat{s}e_{\hat{\theta}^*} \right],
\]  

(3.31)

Normal confidence interval

\(z(\alpha)\)  
Percentage point of normal distribution

Example

\(z(1 - 0.025) \approx 1.959964\)

“±2-sigma interval is 95% Cl.”
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

Student’s t confidence interval

\[ CI_{\hat{\theta}, 1-2\alpha} = \left[ \hat{\theta} + t_{\nu}(\alpha) \cdot \hat{s}e_{\hat{\theta}^*}; \hat{\theta} - t_{\nu}(\alpha) \cdot \hat{s}e_{\hat{\theta}^*} \right], \quad (3.32) \]

**$t_{\nu}(\alpha)$**  
Percentage point of Student’s $t$-distribution  
with $\nu$ degrees of freedom ($\nu = n – \text{number of parameters}$)  
Takes into account that not only standard error,  
but also fit value estimated.  
Negligible difference to normal CI if $\nu$ above ~ 30
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

\begin{align*}
\text{Sample, estimate} & \quad \{t(i), x(i)\}_{i=1}^n \\
\text{Replications} & \quad \{t^{*1}(i), x^{*1}(i)\}_{i=1}^n, \quad \{t^{*2}(i), x^{*2}(i)\}_{i=1}^n, \quad \{t^{*B}(i), x^{*B}(i)\}_{i=1}^n \\
\text{Confidence interval} & \quad \text{CI}_{\hat{\theta}, 1-2\alpha} \\
\hat{\theta}^{*1} & \quad \hat{\theta}^{*2} \quad \hat{\theta}^{*B} \\
\end{align*}
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

Construction of equi-tailed percentile CIs

1. $\hat{\theta}, \{\hat{\theta}^*_b\}_{b=1}^B$

2. Percentile CI

5%  90%  5%
3. Bootstrap Confidence Intervals

3.4 Bootstrap confidence intervals

Construction of equi-tailed percentile CIs

1. $\hat{\theta}, \{\hat{\theta}^*_{b}\}_{b=1}^B$

2. Percentile CI

3. $\hat{\theta}, \text{med}\{\hat{\theta}^*_{b}\}_{b=1}$

4. Bias correction

5% 90% 5%
3. Bootstrap Confidence Intervals

3.6 Bootstrap hypothesis tests

Statistical inference

- Estimating parameter $\theta$
- Testing hypothesis (true/false)
3. Bootstrap Confidence Intervals

3.6 Bootstrap hypothesis tests

$H_0$  Null hypothesis
$H_1$  Alternative hypothesis ($H_0$ and $H_1$ mutually exclusive)
$U, u$  Test statistic

Our choice, a guide to help us decide ($H_0$ or $H_1$)

$F_0(u)$  Distribution function of $U$ “under $H_0$”

$$P = \text{prob}(U \geq u \mid H_0)$$
$$= 1 - F_0(u).$$

$P$  $P$-value (one-sided)

$$P = \text{prob}(|U| \geq |u| \mid H_0).$$

$P$-value (two-sided)

If $P$ is small, then we reject $H_0$ and accept $H_1$. 

Book, Eqs. (3.39), (3.40)
3. Bootstrap Confidence Intervals

3.6 Bootstrap hypothesis tests

$H_0$  Null hypothesis ("Gaussian white noise process, mean $\mu = 0$")

$H_1$  Alternative hypothesis ($H_0$ and $H_1$ mutually exclusive) ("$\mu \neq 0$")

$U, u$  Test statistic (sample mean, $U = \Sigma X(i)/n, u = \Sigma x(i)/n$)

Our choice, a guide to help us decide ($H_0$ or $H_1$)

$F_0(u)$  Distribution function of $U$ "under $H_0$" (Student’s t-distribution)

$$P = \text{prob}(U \geq u \mid H_0)$$
$$= 1 - F_0(u).$$

$P$  $P$-value (one-sided)

$$P = \text{prob}(|U| \geq |u| \mid H_0).$$

$P$-value (two-sided)

If $P$ is small, then we reject $H_0$ and accept $H_1$. 

Book, Eqs. (3.39), (3.40)
3. Bootstrap Confidence Intervals

3.6 Bootstrap hypothesis tests

Real world

Complex processes (distributional shape, persistence, spacing)
Complex constructions of test statistic $u$ (e.g., fingerprint test)
$F_0(u)$ cannot be derived analytically

Bootstrap

Real world
Method to approximate $F_0(u)$
Give \( P \)-values (do not just state whether e.g. \( P < 0.01 \)).

A tested hypothesis should be realistic, not a “straw man”!

A confidence interval contains more quantitative information than a hypothesis test result.
4. Regression I

4.1 Linear regression
4.2 Nonlinear regression
4.3
4. Regression I

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \quad (1.2) \]

Climate equation, stochastic process \( X(i) \)

Time series \( \{t(i), x(i)\}_{i=1}^{n} \)

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

Ignore \( X_{\text{out}}(i) \),
focus on \( X_{\text{trend}}(i) \) and also \( S(i) \).
4. Regression I

$X_{\text{trend}}(i)$

Parametric trend model
  - Linear
  - Nonlinear

Nonparametric trend model
4. Regression I

4.1 Linear regression

\[ X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \] (4.3)

Linear trend model

\[ SSQW(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[ x(i) - \beta_0 - \beta_1 t(i) \right]^2 / S(i)^2, \] (4.4)

Estimation

Weighted least squares (WLS)
where the climate equation without outlier component is then written in dis-

\[ \hat{\beta}_0 = \left[ \sum_{i=1}^{n} x(i)/S(i)^2 - \hat{\beta}_1 \sum_{i=1}^{n} t(i)/S(i)^2 \right] / W, \]  

\hspace{1cm} (4.5)

\[ \hat{\beta}_1 = \left\{ \left[ \sum_{i=1}^{n} t(i)/S(i)^2 \right] \left[ \sum_{i=1}^{n} x(i)/S(i)^2 \right] / W - \sum_{i=1}^{n} t(i) x(i)/S(i)^2 \right\} 
\times \left\{ \left[ \sum_{i=1}^{n} t(i)/S(i)^2 \right]^2 / W - \sum_{i=1}^{n} t(i)^2/S(i)^2 \right\}^{-1}, \]  

\hspace{1cm} (4.6)

\[ W = \sum_{i=1}^{n} 1/S(i)^2. \]  

\hspace{1cm} (4.7)
4. Regression I

4.1 Linear regression

\[ X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \]  

(4.3)

Linear trend model

\[ SSQW(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{[x(i) - \beta_0 - \beta_1 t(i)]^2}{S(i)^2}, \]  

(4.4)

Estimation

Ordinary least squares (OLS)

\[ S(i) = S = \text{const. (known or unknown)} \]

no weighting, even easier to calculate than WLS
4. Regression I

4.1 Linear regression

Figure 4.1. Linear regression models fitted to modelled Arctic river runoff (Fig. 1.9).

a Natural forcing only; b combined anthropogenic and natural forcing. Following $\hat{\beta}_0 = 3068 \text{ km}^3\text{a}^{-1}$, $\hat{\beta}_1 = 0.102 \text{ km}^3\text{a}^{-2}$
4. Regression I

4.1 Linear regression

\[ X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \]  \hspace{1cm} (4.3)

Linear trend model

\[ SSQG(\beta) = (x - T\beta)' V^{-1} (x - T\beta). \]  \hspace{1cm} (4.9)

Estimation

Generalized least squares (GLS)

\[ V, \text{ covariance matrix (contains persistence and variability) } \]
4. Regression I

4.1 Linear regression

\[ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ (parameter vector),} \quad (4.10) \]

\[ x = \begin{bmatrix} x(1) \\ \vdots \\ x(n) \end{bmatrix} \text{ (data vector),} \quad (4.11) \]

\[ T = \begin{bmatrix} 1 & t(1) \\ \vdots & \vdots \\ 1 & t(n) \end{bmatrix} \text{ (time matrix)} \quad (4.12) \]
4. Regression I

4.1 Linear regression

\[ \hat{\beta} = (T'V^{-1}T)^{-1} T'V^{-1}x. \] (4.13)

\[ V(i_1, i_2) = S(i_1) \cdot S(i_2) \cdot E[X_{\text{noise}}(i_1) \cdot X_{\text{noise}}(i_2)], \] (4.14)

Covariance matrix elements

\[ \hat{V}(i_1, i_2) = \hat{S}(i_1) \cdot \hat{S}(i_2) \cdot \exp \left[ -|t(i_1) - t(i_2)| / \hat{\tau}' \right], \] (4.15)

Covariance matrix elements

estimated

AR(1) persistence

bias-corrected persistence time
4. Regression I

4.1 Linear regression

Regression residuals, weighted

\[ r(i) = \frac{x(i) - \hat{\beta}_0 - \hat{\beta}_1 t(i)}{\hat{S}(i)}, \quad (4.16) \]

Regression residuals, unweighted

\[ e(i) = \frac{x(i) - \hat{\beta}_0 - \hat{\beta}_1 t(i)}{\hat{S}(i)} \]
4. Regression I

4.1 Linear regression

Assume

1. Noise has Gaussian shape
2. Covariance matrix known (or correctly estimated)
3. Linear regression model suitable

Then

\[
\text{CI}_{\beta_j, 1-2\alpha} = \left[ \hat{\beta}_j + t_{n-2}(\alpha) \cdot \text{se}_{\hat{\beta}_j}; \hat{\beta}_j + t_{n-2}(1 - \alpha) \cdot \text{se}_{\hat{\beta}_j} \right],
\]

Classical CI \((j = 0, \text{intercept}; j = 1, \text{slope})\)

\[
\text{se}_{\hat{\beta}_j} = [C(j, j)]^{1/2}, \quad \text{C} = (T'V^{-1}T)^{-1}.
\]
4. Regression I
4.1 Linear regression

OLS

special case of GLS

zero off-diagonal and equal $V$ diagonal matrix elements

\[
\text{se} \hat{\beta}_0 = MS_E^{1/2} \left\{ \frac{1}{n} + \left( \sum_{i=1}^{n} t(i)/n \right)^2 / \left[ \sum_{i=1}^{n} t(i)^2 - \left( \sum_{i=1}^{n} t(i) \right)^2 / n \right] \right\}^{1/2},
\]

(4.24)

\[
\text{se} \hat{\beta}_1 = MS_E^{1/2} \left[ \sum_{i=1}^{n} t(i)^2 - \left( \sum_{i=1}^{n} t(i) \right)^2 / n \right]^{-1/2}.
\]

(4.25)

\[
\widehat{S}(i) = \widehat{S} = \left\{ \sum_{i=1}^{n} \left[ x(i) - \hat{\beta}_0 - \hat{\beta}_1 t(i) \right]^2 / (n - 2) \right\}^{1/2} = MS_E^{1/2}.
\]

(4.8)

$MS_E$, residual mean square
4. Regression I
4.1 Linear regression

Conclusion 1

If we know

variability $S(i)$
persistence

$\tau$ (uneven spacing)
a (even spacing)

(that means, covariance matrix $\mathbf{V}$)

$\Rightarrow$ Simple calculations, straightforward estimation
4. Regression I

4.1 Linear regression

Conclusion 2

\[ \text{OLS} \in \text{WLS} \in \text{GLS} \]
4. Regression I

4.1 Linear regression

Real world

- unknown variability \( S(i) \)
- unknown persistence
  - \( \tau \) (uneven spacing)
  - \( a \) (even spacing)
- that means, unknown covariance matrix \( V \)

⇒ Iteratively (Prais–Winsten procedure)
⇒ Quick-and-dirty: OLS with effective data size
4. Regression I

4.1 Linear regression

Via effective data size, \( n'_\mu < n \)

\[
\hat{S}(i) = \hat{S} = \left\{ \sum_{i=1}^{n} \left[ x(i) - \hat{\beta}_0 - \hat{\beta}_1 t(i) \right]^2 / (n - 2) \right\}^{1/2} = MS_E^{1/2}. \tag{4.8}
\]

Insert in Eq. (4.8) \( n'_\mu \) for \( n \)

⇒ Larger residual mean square

⇒ Wider classical OLS CI

Caveats

No re-estimation of persistence parameter

\( n'_\mu \) applies to mean estimation, not regression
4. Regression I

4.1 Linear regression

Table 4.1. Monte Carlo experiment, linear OLS regression with AR(1) noise of normal shape, even spacing: CI coverage performance. \( n_{\text{sim}} = 47,500 \) random samples were generated from \( X(i) = 2 + 2T(i) + X_{\text{noise}}(i), \) where \( T(i) = i, i = 1, \ldots, n \) and the noise is a Gaussian AR(1) process (Eq. 2.1) with \( a = 1/e \approx 0.37. \) Two CI types for the estimated slope were constructed, classical and bootstrap. Construction of classical CIs either ignored persistence and calculated via \( n \) (Eqs. 4.8, 4.20 and 4.25) or used \( n'_{\mu} \) (Section 4.1.4.3). The bootstrap CIs used ARB (Algorithm 3.4) or MBB (Algorithm 3.1) resampling and the BCa method (Section 3.4.4) with \( B = 1999 \) and \( \alpha = 0.025. \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \gamma_{\beta_1}^a )</th>
<th>( \gamma_{\beta_1}^{n'_{\mu}} )</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classical Via ( n ) Via ( n'_{\mu} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.851</td>
<td>0.900</td>
<td>0.950</td>
</tr>
<tr>
<td>20</td>
<td>0.832</td>
<td>0.915</td>
<td>0.950</td>
</tr>
<tr>
<td>50</td>
<td>0.819</td>
<td>0.932</td>
<td>0.950</td>
</tr>
<tr>
<td>100</td>
<td>0.817</td>
<td>0.941</td>
<td>0.950</td>
</tr>
<tr>
<td>200</td>
<td>0.819</td>
<td>0.945</td>
<td>0.950</td>
</tr>
<tr>
<td>500</td>
<td>0.818</td>
<td>0.947</td>
<td>0.950</td>
</tr>
<tr>
<td>1000</td>
<td>0.816</td>
<td>0.950</td>
<td>0.950</td>
</tr>
</tbody>
</table>

Gaussian shape (Table 4.1). Ignoring persistence leads to underestimating the \( \text{MS}\text{E} \) and standard errors and yields therefore too narrow CIs (Table 4.2). Noise of AR(1) persistence and Gaussian shape is rather easy to handle for the remaining three CI types. In particular, classical CIs via \( n \) and bootstrap BCa CIs with ARB resampling performed well in terms of coverage accuracy (Table 4.1); the classical CIs did even better than the bootstrap CIs for small data sizes (less than 100).

Retaining the AR(1) persistence model but going from Gaussian to lognormal distributional shape diminished only slightly the coverage performance (Table 4.3). Interestingly, the classical CI via \( n \) performed also here excellently already for small data sizes.

Retaining the Gaussian distributional shape but adopting a persistence model more complex than AR(1) had more severe effects on coverage performance than changing the shape. In the case of AR(2) persistence (Table 4.4), the classical CI via \( n \) as well as the bootstrap BCa CI with ARB resampling failed for all data sizes tested. The reason is that the AR(1) assumption, made for \( n \) calculation as well as ARB
4. Regression I

4.1 Linear regression

Figure 4.1. Linear regression models fitted to modelled Arctic river runoff (Fig. 1.9). a Natural forcing only; b combined anthropogenic and natural forcing. Following Wu et al. (2005), the fits (solid lines) were obtained by OLS regression using the data from (a) the whole interval 1900–1996 and (b) from two intervals, 1936–2001 and 1965–2001. The estimated regression parameters (Eqs. 4.5 and 4.6) and their standard errors (Eqs. 4.24 and 4.25) are as follows.

\begin{align*}
\hat{\beta}_0 &= 3068 \pm 694 \text{ km}^3\text{a}^{-1}, \quad \hat{\beta}_1 = 0.102 \pm 0.356 \text{ km}^3\text{a}^{-2} \text{ (via } n; \text{ Wu et al. 2005)} \\
\hat{\beta}_0 &= 3068 \pm 859 \text{ km}^3\text{a}^{-1}, \quad \hat{\beta}_1 = 0.102 \pm 0.441 \text{ km}^3\text{a}^{-2} \text{ (via } n').
\end{align*}

4.1.2 Generalized least-squares estimation

In a practical climatological setting, noise \((i)\) often exhibits persistence. This means more structure or information content than a purely random process has. This knowledge can be used to apply the generalized least-squares (GLS) estimation, where the following sum of squares is minimized:

\[
SSQ_G(\beta) = (x - \beta)^T V^{-1} (x - \beta).
\]
4. Regression I

4.2 Nonlinear regression

\[ X(i) = \beta_0 + \beta_1 T(i) + \beta_2 T(i)^2 + S(i) \cdot X_{\text{noise}}(i), \] (4.37)

Parabolic trend model: not really nonlinear (i.e, in parameters)

\[ X(i) = \beta_0 \{1 - \exp[-\beta_1 T(i)]\} + S(i) \cdot X_{\text{noise}}(i). \] (4.38)

Saturation trend function: nonlinear
4. Regression I
4.2 Nonlinear regression

High-dimensional parameter space ($\beta$)

Estimation
Search for minimum of $SSQG(\beta)$

Often no exact solution,
but numerical techniques
(1) Guess starting values.
(2) Go step (of defined size) into negative gradient direction.
(3) If changes small: exit, otherwise go back to (1).
4. Regression I

4.2 Nonlinear regression

Trend-change model: break

\[ X_{\text{break}}(T) = \begin{cases} 
  x_1 + (T - t_1)(x_2 - x_1)/(t_2 - t_1) & \text{for } T \leq t_2, \\
  x_2 + (T - t_2)(x_3 - x_2)/(t_3 - t_2) & \text{for } T > t_2, 
\end{cases} \] (4.47)
4. Regression I

4.2 Nonlinear regression

Trend-change model: break

\[ SSQW(x_1, t_2, x_2, x_3) = \sum_{i=1}^{n} \frac{[x(i) - x_{\text{break}}(i)]^2}{S(i)^2}, \quad (4.48) \]

Estimation (WLS combined with brute-force search for \( t_2 \))
4. Regression I

4.2 Nonlinear regression

Break point in Arctic river runoff

\[ \hat{t}_2 = 1973 \]

\[ \hat{\beta}_1 = -1.8 \text{ km}^3\text{a}^{-2} \]

\[ \hat{\beta}_2 = 9.7 \text{ km}^3\text{a}^{-2} \]
4. Regression I

4.2 Nonlinear regression

Error bars?

Classical formulas do not exist, gradient not defined

⇒ Bootstrap resampling

Break point in Arctic river runoff
\[ t_2 = 1973 \]
\[ \hat{\beta}_1 = -1.8 \text{ km}^3\text{a}^{-2} \]
\[ \hat{\beta}_2 = 9.7 \text{ km}^3\text{a}^{-2} \]
4. Regression I

4.2 Nonlinear regression

\[ e(i) = x(i) - \hat{x}_{\text{break}}(i), \ i = 1, \ldots, n; \]

Unweighted residuals

\[ r(i) = e(i) / \hat{S}(i), \ i = 1, \ldots, n. \]

Weighted residuals
4. Regression I

4.2 Nonlinear regression

Bootstrap error bars
Take fit curve
Add blocks of residuals
Re-estimate: replication
Repeat many times
Calculate standard deviation (replications)
Break point in Arctic river runoff

\[ \hat{t}_2 = 1973 \pm 6 \]

\[ \hat{\beta}_1 = -1.8 \pm 1.6 \text{ km}^3\text{a}^{-2} \]

\[ \hat{\beta}_2 = 9.7 \pm 3.6 \text{ km}^3\text{a}^{-2} \]
4. Regression I

4.2 Nonlinear regression—**Homework**

![Break regression model](image1)

![Ramp regression model](image2)

---

Solomon et al. 2007 (p. 37)

www.manfredmudelsee.com/soft/BREAKFIT

www.manfredmudelsee.com/soft/RAMPFIT
(1) \[ \hat{\theta} \neq \theta \]

(2) Computational techniques (bootstrap, Monte Carlo) help to obtain realistic error bars.

(3) Estimation problems (linear, break, ramp, etc.): Be creative!
Climate Time Series Analysis: Conclusions

(1) $\hat{\theta} \neq \theta$

(2) Computational techniques (bootstrap, Monte Carlo) help to obtain realistic error bars.

(3) Estimation problems (linear, break, ramp, etc.): Be creative!

Thanks!

Postdoc job: www.climate-risk-analysis.com
Climate, $X(T)$
Weather, $Y(T)$

\[
\frac{dX(T)}{dT} = F(X(T), Y(T)), \quad \text{timescale } \tau_X,
\]

\[
\frac{dY(T)}{dT} = G(X(T), Y(T)), \quad \text{timescale } \tau_Y,
\]

Weather components
“leave a trace” in climate.

\[
\frac{dX(T)}{dT} \simeq F(X(0), Y(T)),
\]

\[
= W(T),
\]

\[
X(T + 1) = X(T) + \mathcal{E}_{N(0, \sigma^2)}(T).
\]

Climate does not run away:
negative feedback ($\sim X$)

\[
X(T + 1) = a \cdot X(T) + \mathcal{E}_{N(0, \sigma^2)}(T),
\]

$W(T)$, Wiener process (“continuous-time noise increment”)
Hasselman (1976); Book, Eqs. (2.25), (2.26), (2.28), (2.29), (2.30), Section 2.5.1
2. Persistence Models

2.5 Climate theory

Support of simple AR(1) climate noise model

1. Weather components (Hasselmann 1976)

2. Climate equation

\[ X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \]

3. We have also trend, outliers/extremes and variability to describe climate.
3. Bootstrap Confidence Intervals

3.3 Bootstrap resampling

Block length selector (MBB) after Carlstein (1986)

\[
\text{l}_{\text{opt}} = \text{NINT}\left\{ \left[ 6^{1/2} \cdot \hat{a} / \left(1 - \hat{a}^2\right) \right]^{2/3} \cdot n^{1/3} \right\}, \quad (3.28)
\]

Other block length selectors via
- persistence time
- autocorrelation function
4. Regression I

4.1 Linear regression

Step 1  Make an initial guess of the variability, $\hat{S}(i)$. Make an initial guess of the persistence time, $\hat{\tau}$, for uneven spacing or the autocorrelation parameter, $\hat{a}$, for even spacing. (For notational convenience, we omit writing a superscript.) Often the simple choices of constant variability and absent persistence are sufficient.

Step 2  Update the $V$ matrix after Eq. (4.15).

Step 3  Perform GLS (Eq. 4.13) to obtain updates of regression parameter estimates. Update also $SSQG$ (Eq. 4.9).

Step 4  If regression parameters, and possibly also $SSQG$, have not changed strongly at the preceding step, take this solution and go to Step 8.

Step 5  Calculate the unweighted regression residuals $e(i)$. Update the variability estimate $\hat{S}(i)$ by using the $e(i)$.

Step 6  Calculate the weighted regression residuals, $r(i)$, after Eq. (4.16). Update the persistence estimate $\hat{\tau}$ or $\hat{a}$ by using the $r(i)$. Take bias correction (Section 2.6) into account.

Step 7  Go to Step 2.
Step 7 Go to Step 2.

Step 8 After regression parameters, and possibly also SSQG, approached the solution at Step 4, calculate the residuals $r(i)$ and the standard errors (Eq. 4.21). The residuals can be used for graphical analysis (Montgomery and Peck 1992). One looks whether they fail to reveal more structure than a Gaussian AR(1) process, that means, whether the linear regression model with AR(1) noise is a suitable description of the data. The standard errors can be used for classical CI construction (Eq. 4.20).