Time series in epidemiology

Emanuele Giorgi

CHICAS, Lancaster Medical School, Lancaster University, UK

Objectives of this module

- To understand the limitations of standard linear regressions models for time series data.
- To model temporal trends (both seasonal and non-seaonal) through the use of explanatory variables.
- To understand and apply basic models for time series analysis.

1. Review of linear regression

Linear regression model

$$Y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + Z_i : i = 1, ..., n$$

- explanatory variables/factors $x_{1i}, ..., x_{ki}$
- regression parameters $\beta_0, ..., \beta_k$
- residuals Z_i
- responses Y_i

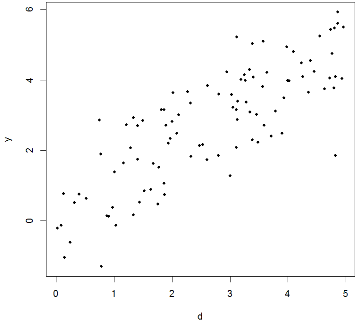
```
fit1<-lm(y x); summary(fit1)
xsq<-x*x; fit2<-lm(y x+xsq); summary(fit2)
names(fit2)
xx<-0.1*(0:200); beta<-fit2$coef;
ff<-beta[1]+beta[2]*xx+beta[3]*xx*xx
par(mfrow=c(1,1)); plot(x,y,pch=19,xlim=c(0,20));
lines(xx,ff,col="blue",lwd=3)
```

Parameter estimation: simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + Z_i : i = 1, ..., n$$

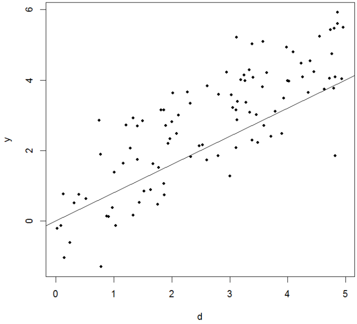
Parameter estimation: simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + Z_i : i = 1, ..., n$$



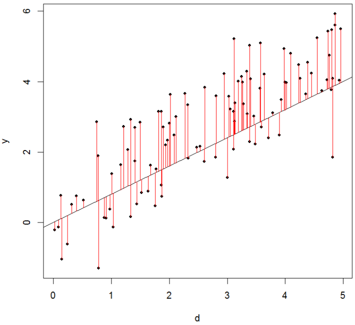
Parameter estimation: simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + Z_i : i = 1, ..., n$$



Parameter estimation: simple linear regression

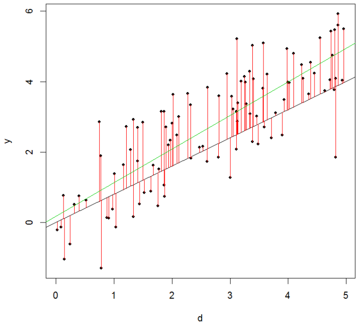
$$Y_i = \beta_0 + \beta_1 x_i + Z_i : i = 1, ..., n$$



$$RSS = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$$

Parameter estimation: simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + Z_i : i = 1, ..., n$$



$$RSS = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$$

General linear model: things to remember

- Transformation of x and/or Y widens the scope of the model

  Examples:
  - $Y_i = \beta_1 + \beta_2 \log x_i + Z_i$
  - $\log Y_i = \beta_1 + \beta_2 x_i + Z_i$
- Normality of the Z_i is less important than independence and constant variance
- But if the Z_i are iid $\mathrm{N}(0, \sigma^2)$, likelihood-based inference is straightforward (and least squares estimates are maximum likelihood estimates)
- Diagnostic checking of residuals is an important part of model-building, but requires subjective judgement.

Discrete time series

Discrete time series

- t = time of observation

Discrete time series

- t = time of observation
- Y_t = random variable associated with **discrete time-series**.

Discrete time series

- t = time of observation
- Y_t = random variable associated with **discrete time-series**.
- Time matters

$$\mathrm{Cov}\{Y_t, Y_{t-k}\} \neq 0, k = 0, \pm 1, \pm 2, \ldots$$

Discrete time series

- t = time of observation
- Y_t = random variable associated with **discrete time-series**.
- Time matters

$$\mathrm{Cov}\{Y_t, Y_{t-k}\} \neq 0, k = 0, \pm 1, \pm 2, \ldots$$

- $\mathrm{Y}^\top = (Y_1, \ldots, Y_t)$

Discrete time series

- t = time of observation
- Y_t = random variable associated with **discrete time-series**.
- Time matters

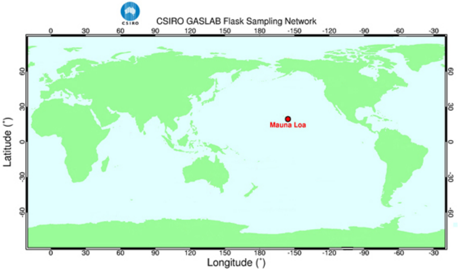
$$\mathrm{Cov}\{Y_t, Y_{t-k}\} \neq 0, k = 0, \pm 1, \pm 2, \ldots$$

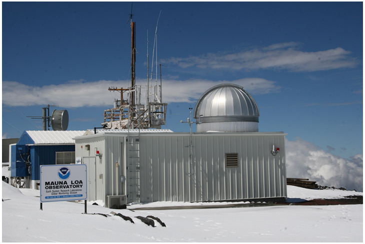
- $\mathrm{Y}^\top = (Y_1, \ldots, Y_t)$
- Our model for nature,

$$[\mathrm{Y}] = [Y_1] \times [Y_2|Y_1] \times \ldots \times [Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_1].$$

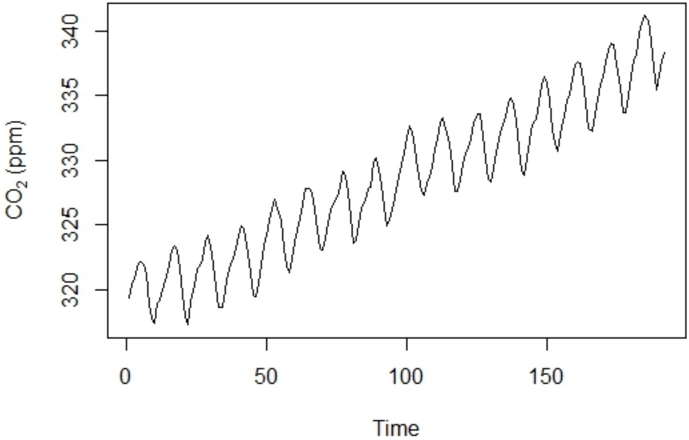
Example: CO_2 time series

http://yearsoflivingdangerously.com/watch/season-1/

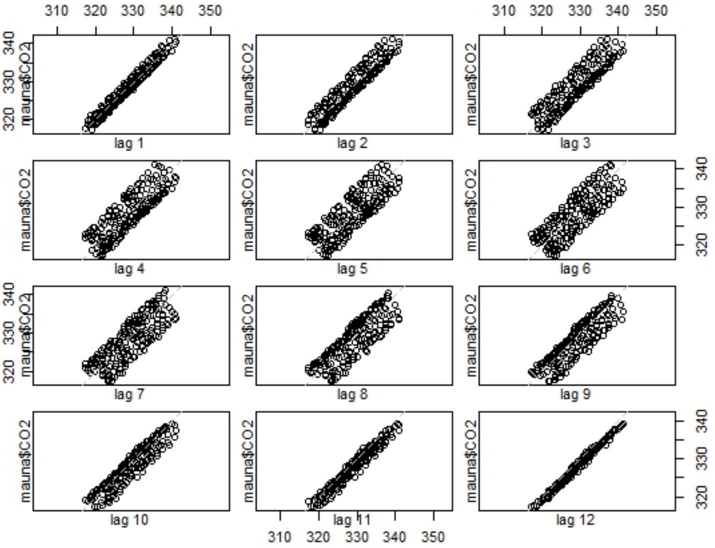




Example: CO_2 time series 



Lag plots 



310 320 330 340 350
mauna$CO2
320 330 340
lag 1
lag 2
lag 3
lag 4
lag 5
lag 6
lag 7
lag 8
lag 9
lag 10
lag 11
lag 12

Harmonic regression

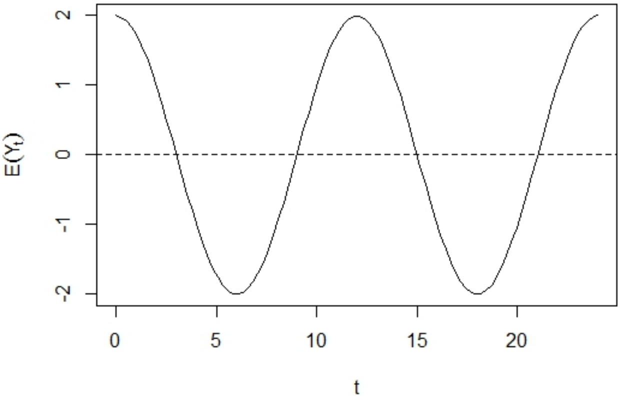
$$E[Y_t] = \beta_1 + A \sin\{2\pi f t + \phi\}$$

Harmonic regression

$$\begin{aligned} E[Y_t] &= \beta_1 + A\sin\{2\pi ft + \phi\} \\ &= \beta_1 + \beta_2\sin(2\pi ft) + \beta_2\cos(2\pi ft). \end{aligned}$$

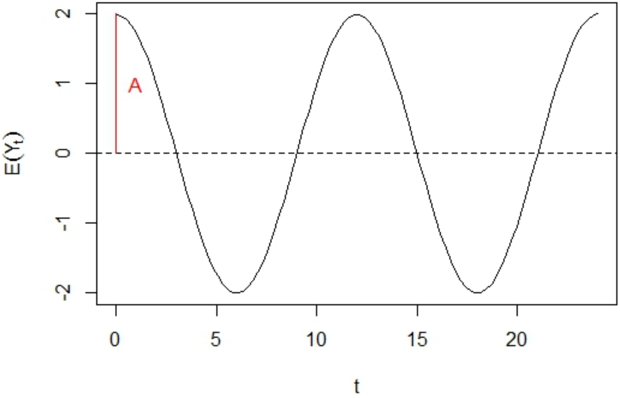
Harmonic regression

$$\begin{aligned} E[Y_t] &= \beta_1 + A\sin\{2\pi ft + \phi\} \\ &= \beta_1 + \beta_2 \sin(2\pi ft) + \beta_2 \cos(2\pi ft). \end{aligned}$$



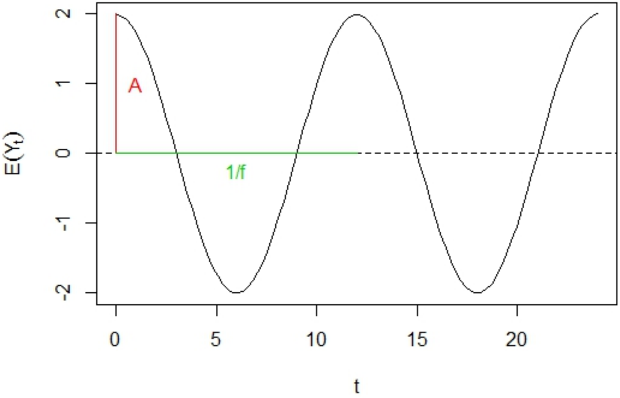
Harmonic regression

$$
\begin{aligned}
E[Y_t] &= \beta_1 + A \sin\{2\pi f t + \phi\} \\
&= \beta_1 + \beta_2 \sin(2\pi f t) + \beta_2 \cos(2\pi f t).
\end{aligned}
$$

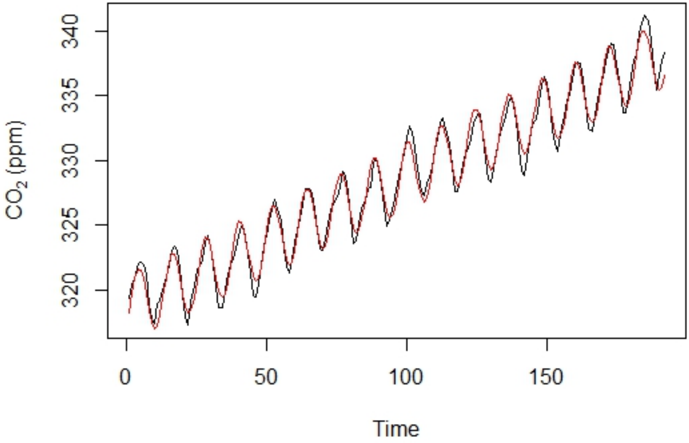


Harmonic regression

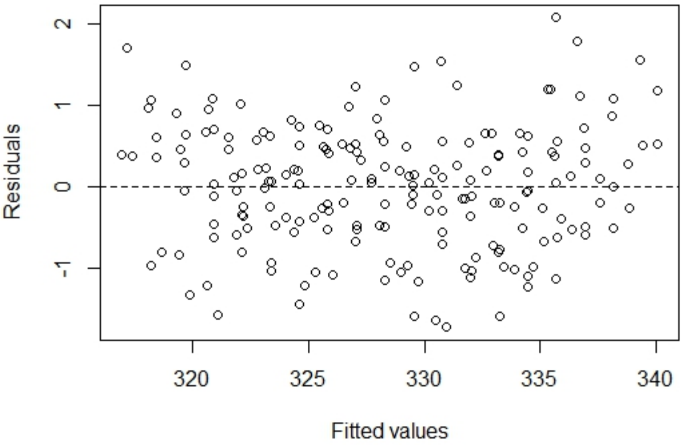
$$\begin{aligned} E[Y_t] &= \beta_1 + A\sin\{2\pi ft + \phi\} \\ &= \beta_1 + \beta_2\sin(2\pi ft) + \beta_2\cos(2\pi ft). \end{aligned}$$



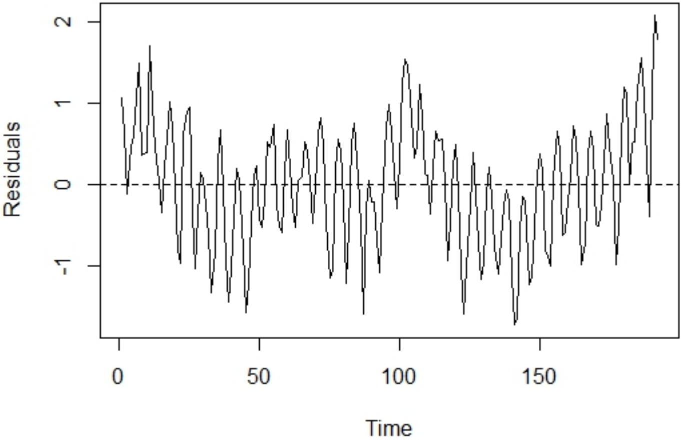
Do what you know from the linear model 



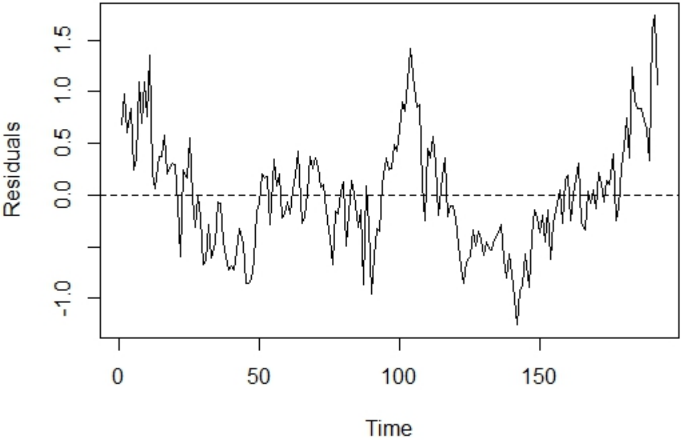
Do what you know from the linear model 



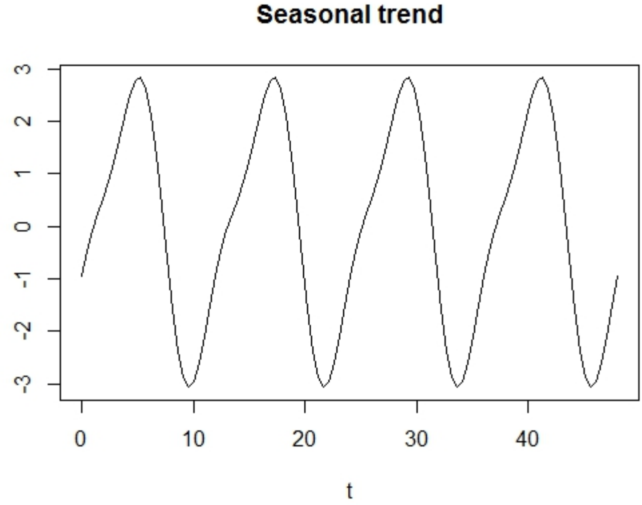
Do what you know from the linear model 



Do what you know from the linear model 



Do what you know from the linear model



Stationarity

Stationarity

- First order stationarity

$$E[Y_t] = \mu, \text{ for all } t.$$

Stationarity

- First order stationarity

$$E[Y_t] = \mu, \text{ for all } t.$$

- Second order stationarity

$$V[Y_t] = \sigma^2, \text{ for all } t$$

and

$$\mathrm{COV}(Y_t, Y_{t-k}) = \gamma_k = \sigma^2 \rho_k.$$

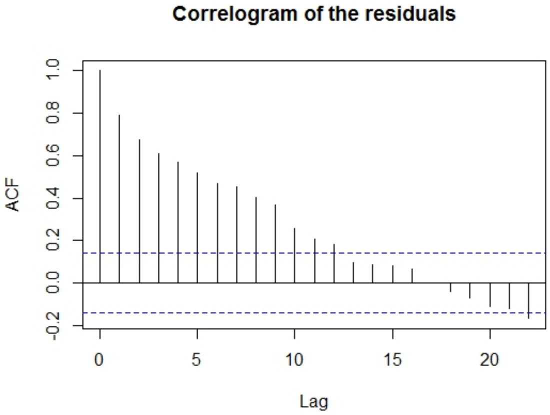
- What value does γ_0 take?

Correlogram 

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

Correlogram 

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t-k} - \bar{y})$$



Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- $E[Y_t] = 0$ (without loss of generalization).

Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- $E[Y_t] = 0$ (without loss of generalization).
- Lag operator: $BY_t = Y_{t-1}$.

Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- $E[Y_t] = 0$ (without loss of generalization).
- Lag operator: $BY_t = Y_{t-1}$.
- $B^k Y_t = Y_{t-k}$.

Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- $E[Y_t] = 0$ (without loss of generalization).
- Lag operator: $BY_t = Y_{t-1}$.
- $B^k Y_t = Y_{t-k}$.
- $Y_t = \sum_{k=1}^{p} \phi_k B^k Y_t + Z_t \iff$
  $(1 - \phi_1 B - \phi_2 B^2 - \cdots - B^p \phi_p) Y_t = Z_t$

Autoregressive models

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t.$$

- $E[Y_t] = 0$ (without loss of generalization).
- Lag operator: $BY_t = Y_{t-1}$.
- $B^k Y_t = Y_{t-k}$.
- $Y_t = \sum_{k=1}^{p} \phi_k B^k Y_t + Z_t \iff$
  $(1 - \phi_1 B - \phi_2 B^2 - \cdots - B^p \phi_p) Y_t = Z_t$

Stationarity

An AR(p) process is stationary if and only if the roots of the equation

$$1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p = 0$$

are in absolute value smaller than 1.

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

- $V[Z_t] = \sigma^2$.

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

- $V[Z_t] = \sigma^2$.
- $1 - \phi x = 0 \iff x = \phi^{-1}$, hence $|\phi| < 1$.

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

- $V[Z_t] = \sigma^2$.
- $1 - \phi x = 0 \iff x = \phi^{-1}$, hence $|\phi| < 1$.
- $V[Y_t] = \sigma^2/(1-\phi^2)$, $\gamma_k = \sigma^2\phi^k/(1-\phi^2)$.

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

- $V[Z_t] = \sigma^2$.
- $1 - \phi x = 0 \iff x = \phi^{-1}$, hence $|\phi| < 1$.
- $V[Y_t] = \sigma^2/(1-\phi^2)$, $\gamma_k = \sigma^2\phi^k/(1-\phi^2)$.
- Predictive distribution: $Y_t|Y_{t-1} = y_{t-1} \sim N(\phi y_{t-1}, \sigma^2)$.

Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

- $V[Z_t] = \sigma^2$.
- $1 - \phi x = 0 \iff x = \phi^{-1}$, hence $|\phi| < 1$.
- $V[Y_t] = \sigma^2/(1-\phi^2)$, $\gamma_k = \sigma^2\phi^k/(1-\phi^2)$.
- Predictive distribution: $Y_t | Y_{t-1} = y_{t-1} \sim N(\phi y_{t-1}, \sigma^2)$.
- $\boldsymbol{\theta}^\top = (\phi, \sigma^2)$.

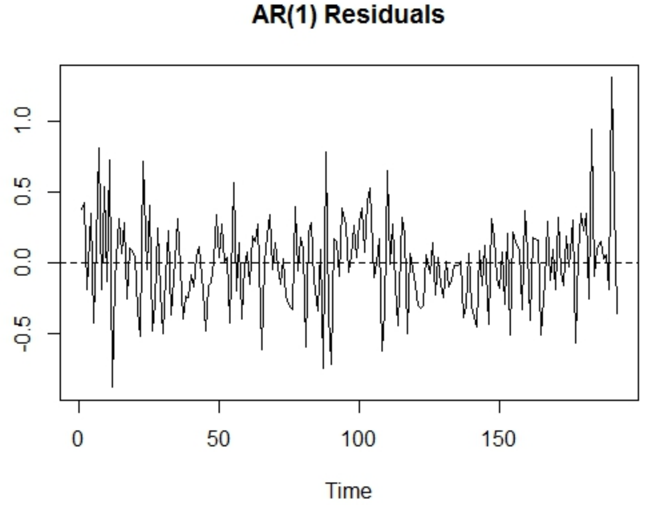
Example: AR(1)

$$Y_t = \phi Y_{t-1} + Z_t$$

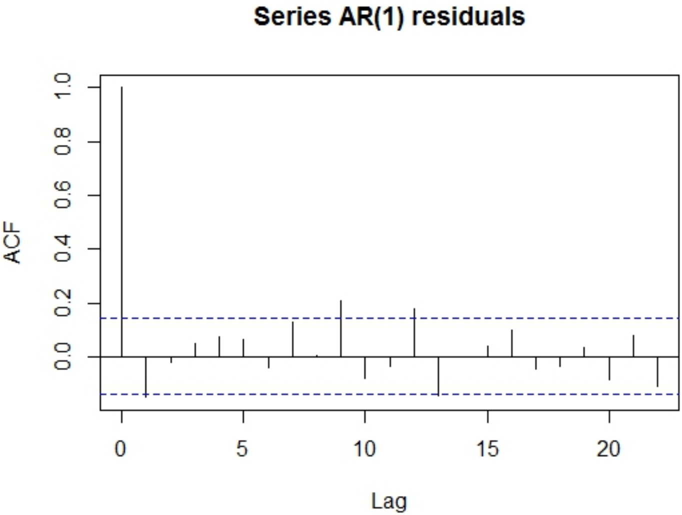
- $V[Z_t] = \sigma^2$.
- $1 - \phi x = 0 \iff x = \phi^{-1}$, hence $|\phi| < 1$.
- $V[Y_t] = \sigma^2/(1-\phi^2)$, $\gamma_k = \sigma^2\phi^k/(1-\phi^2)$.
- Predictive distribution: $Y_t | Y_{t-1} = y_{t-1} \sim N(\phi y_{t-1}, \sigma^2)$.
- $\boldsymbol{\theta}^\top = (\phi, \sigma^2)$.

$$\begin{aligned}\log L(\boldsymbol{\theta}) &= -\frac{1}{2}\Bigg[n \log\{\sigma^2\} - \log(1-\phi^2) + (1-\phi^2)\frac{y_1^2}{\sigma^2} \\ &\qquad + \sum_{t=2}^{n} \frac{(y_t - \phi y_{t-1})^2}{\sigma^2}\Bigg]\end{aligned}$$

CO_2 time series (continue) 



CO_2 time series (continue) 



Continuous time series data

- **Problem:** How do we model an outcome that is not observed at equally spaced time points?

Continuous time series data

- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] $=$ [Covariates effects] $+$ [Temporal random effect] $+$ [Noise]

Continuous time series data

- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] $=$ [Covariates effects] $+$ [Temporal random effect] $+$ [Noise]
- Let t_i, $i = \ldots, n$ denote the observation time points.

Continuous time series data

- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] = [Covariates effects] + [Temporal random effect] + [Noise]
- Let t_i, $i = \ldots, n$ denote the observation time points.
- $W(t)$ is a temporal stochastic process.

Continuous time series data

- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] = [Covariates effects] + [Temporal random effect] + [Noise]
- Let t_i, $i = \ldots, n$ denote the observation time points.
- $W(t)$ is a temporal stochastic process.
- $Z(t)$ is Gaussian noise.

Continuous time series data

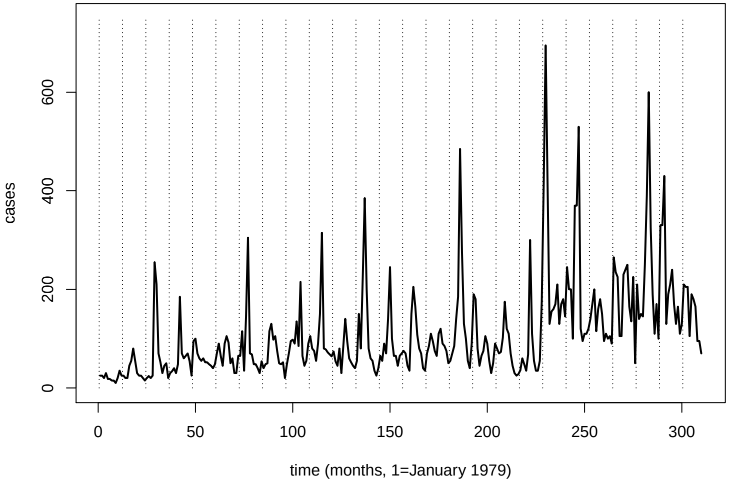
- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] = [Covariates effects] + [Temporal random effect] + [Noise]
- Let t_i, $i = \ldots, n$ denote the observation time points.
- $W(t)$ is a temporal stochastic process.
- $Z(t)$ is Gaussian noise.
- $d(t)$ are temporally referenced covariates.

Continuous time series data

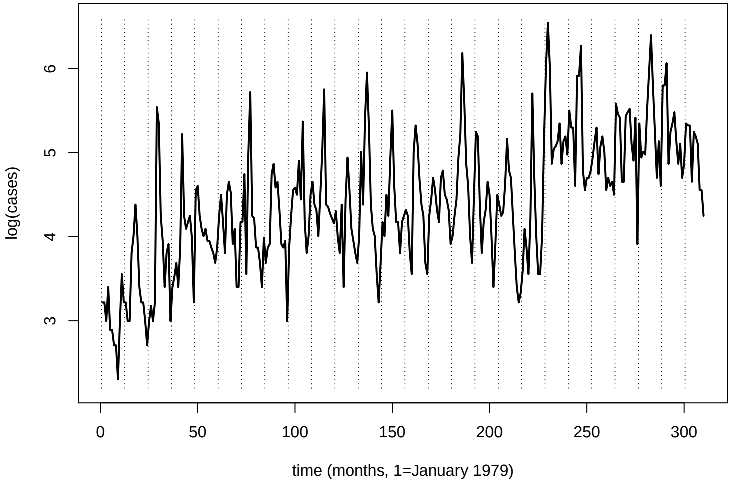
- **Problem:** How do we model an outcome that is not observed at equally spaced time points?
- Continuous time series models: [Outcome] = [Covariates effects] + [Temporal random effect] + [Noise]
- Let t_i, $i = \ldots, n$ denote the observation time points.
- $W(t)$ is a temporal stochastic process.
- $Z(t)$ is Gaussian noise.
- $d(t)$ are temporally referenced covariates.
- A model for the data

$$Y(t_i) = d(t_i)^\top \beta + W(t_i) + Z(t_i).$$

Malaria cases in Kericho, Kenya



Malaria cases in Kericho, Kenya



Standard linear regression analysis

- Objective: accounting for seasonal and non-seasonal trends.

Standard linear regression analysis

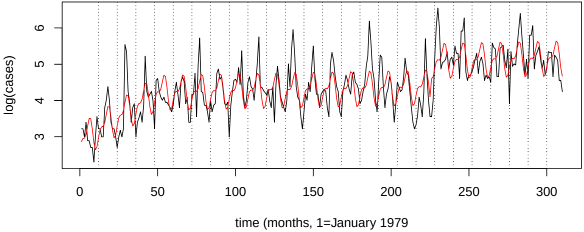
- Objective: accounting for seasonal and non-seasonal trends.
- Let $Y(t)$ denote the log-transformed number of cases.

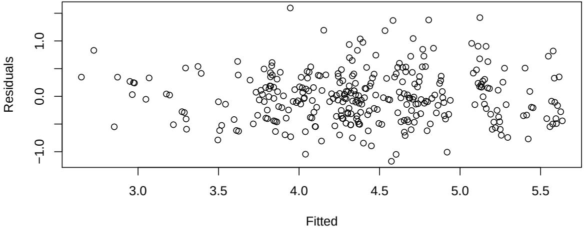
Standard linear regression analysis

- Objective: accounting for seasonal and non-seasonal trends.
- Let $Y(t)$ denote the log-transformed number of cases.
- A model for the data:

$$
\begin{aligned}
Y(t) \quad = \quad & \beta_0 + \beta_1 t + \beta_2 \max\{t - 50, 0\} + \\
& \beta_3 I(t > 225) + \\
& \beta_2 \sin(2\pi t/12) + \beta_3 \cos(2\pi t/12) + \\
& \beta_4 \sin(2\pi t/6) + \beta_5 \cos(2\pi t/6) + \\
& Z(t) \qquad (1)
\end{aligned}
$$

Standard linear regression analysis





The theoretical variogram

The theoretical variogram

- Assumption: $W(t)$ is a stationary zero-mean stochastic process with covariance function

$$\mathrm{Cov}\{W(t), W(t')\} = \sigma^2 \rho(h), h = |t - t'|$$

The theoretical variogram

- Assumption: $W(t)$ is a stationary zero-mean stochatic process with covariance function

$$\mathrm{Cov}\{W(t), W(t')\} = \sigma^2 \rho(h), h = |t - t'|$$

- Assumption: $Z(t)$ are i.i.d. variables with mean 0 and variance τ^2

The theoretical variogram

- Assumption: $W(t)$ is a stationary zero-mean stochastic process with covariance function

$$\mathrm{Cov}\{W(t), W(t')\} = \sigma^2 \rho(h), h = |t - t'|$$

- Assumption: $Z(t)$ are i.i.d. variables with mean 0 and variance τ^2
- Theoretical varioram (definition)

$$\begin{aligned} v(h) &= \frac{1}{2} E[(W(t) + Z(t) - W(t') - Z(t'))^2] \\ &= \tau^2 + \sigma^2 (1 - \rho(h)) \end{aligned}$$

The empirical variogram

The empirical variogram

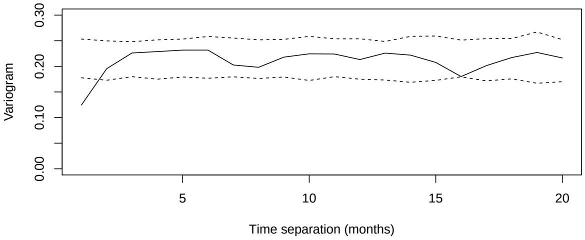
- Let $Z(t)$ denote the residuals from a standard linear regression model.

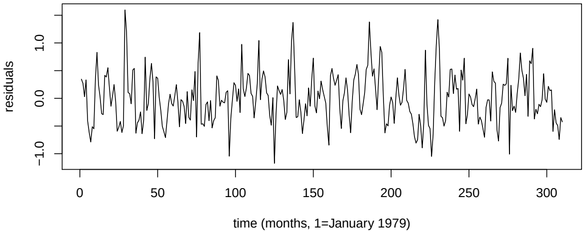
The empirical variogram

- Let $Z(t)$ denote the residuals from a standard linear regression model.
- The empirical variogram (definition)

$$\hat{v}(h) = \frac{1}{2}(\hat{Z}(t) - \hat{Z}(t'))^2, h = |t - t'|. \tag{2}$$

The empirical variogram





$W(t)$: The Matern process

$W(t)$: The Matern process

- Stationary covariance function

$$\operatorname{cov}\{W(t), W(t')\} - \sigma^2 \rho(h), \tag{3}$$

$W(t)$: The Matern process

- Stationary covariance function

$$\mathrm{cov}\{W(t), W(t')\} - \sigma^2 \rho(h), \tag{3}$$

- Matern covariance functions

$$\rho(h) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(h/\phi)^{\kappa}\mathcal{K}_k(h/\phi), h > 0 \tag{4}$$

$W(t)$: The Matern process

- Stationary covariance function

$$\mathrm{cov}\{W(t), W(t')\} - \sigma^2 \rho(h), \tag{3}$$

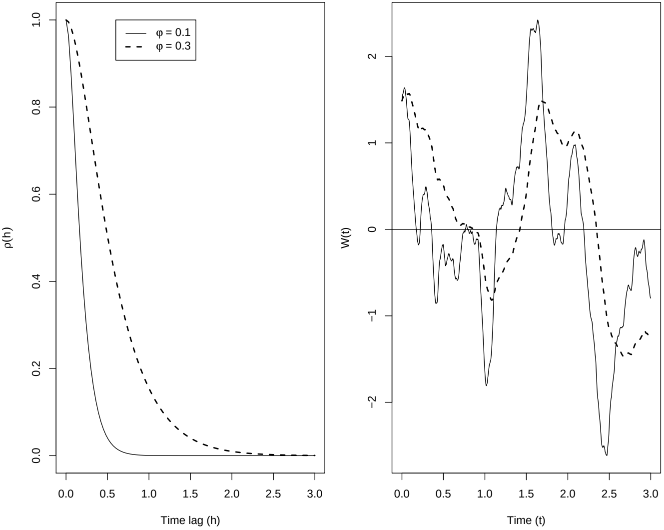
- Matern covariance functions

$$\rho(h) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(h/\phi)^{\kappa}\mathcal{K}_k(h/\phi), h > 0 \tag{4}$$

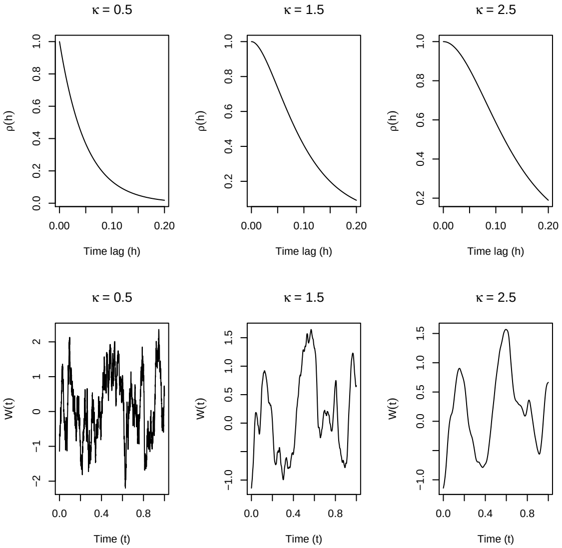
- Special case ($\kappa = 1/2$)

$$\rho(h) = \exp\{-h/\phi\}. \tag{5}$$

The scale parameter ϕ



The smoothness parameter κ



Fitting a Matern process

Fitting a Matern process

- Unkown parameters to estimate: $\theta = (\beta, \sigma^2, \phi, \tau^2)$.

Fitting a Matern process

- Unkown parameters to estimate: $\theta = (\beta, \sigma^2, \phi, \tau^2)$.
- The vector $(Y(t_1), \ldots, Y(t_n))$ follows a multivariate Gaussian distribution with mean $D\beta$ and covariance matrix $\Omega = \sigma^2 R + \tau^2 I$ where

$$[R]_{ij} = \rho(|t_i - t_j|)$$

Fitting a Matern process

- Unkown parameters to estimate: $\theta = (\beta, \sigma^2, \phi, \tau^2)$.
- The vector $(Y(t_1), \ldots, Y(t_n))$ follows a multivariate Gaussian distribution with mean $D\beta$ and covariance matrix $\Omega = \sigma^2 R + \tau^2 I$ where

$$[R]_{ij} = \rho(|t_i - t_j|)$$

- The likelihood function

$$l(\theta) = -\frac{1}{2}\left\{n \log \sigma^2 + \log|\Omega| + (y - D\beta)^\top \Omega^{-1}(y - D\beta)/\sigma^2\right\}, \tag{6}$$

Kericho data: parameter estimation

	$\kappa = 0.5$	
Parameter	Point estimate	95% CI
β_0	2.892	(2.535, 3.248)
β_1	0.137	(0.038, 0.235)
β_2	-0.345	(-0.444, -0.247)
β_3	0.162	(0.086, 0.238)
β_4	0.126	(0.050, 0.202)
β_5	0.026	(0.018, 0.035)
β_6	-0.025	(-0.035, -0.016)
β_7	0.700	(0.397, 1.004)
σ^2	0.214	(0.177, 0.258)
ϕ	1.149	(0.870, 1.516)
τ^2	-	-