# Time Series Analysis

Spring 2024

Peder Bacher and Pernille Yde Nielsen

February 9, 2024

# Outline of the lecture

- Regression based methods, 1st part:
  - Ordinary Least Squares (OLS)
  - Predictions = forecast
  - Global Trend Model
  - Weighted Least Squares (WLS)

# General form of the regression model

$$Y_t = f(\boldsymbol{X}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

# General form of the regression model

$$Y_t = f(\boldsymbol{X}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

$Y_t$ is the output we aim to model

$\boldsymbol{X}_t$ indicates the $p$ independent variables $\boldsymbol{X}_t = (X_{1t}, \cdots, X_{pt})^T$

$t$ is the time index

$\boldsymbol{\theta}$ indicates $m$ unknown parameters $(\theta_1, \cdots, \theta_m)^T$

$\varepsilon_t$ is a sequence of random variables with mean zero, variance $\sigma_t^2$, and $\mathrm{Cov}[\varepsilon_{t_i}, \varepsilon_{t_j}] = \sigma_t^2 \Sigma_{ij}$

# General form of the regression model

$$Y_t = f(\boldsymbol{X}_t, t; \boldsymbol{\theta}) + \varepsilon_t$$

$Y_t$ is the output we aim to model

$\boldsymbol{X}_t$ indicates the $p$ independent variables $\boldsymbol{X}_t = (X_{1t}, \cdots, X_{pt})^T$

$t$ is the time index

$\boldsymbol{\theta}$ indicates $m$ unknown parameters $(\theta_1, \cdots, \theta_m)^T$

$\varepsilon_t$ is a sequence of random variables with mean zero, variance $\sigma_t^2$, and $\mathrm{Cov}[\varepsilon_{t_i}, \varepsilon_{t_j}] = \sigma_t^2 \Sigma_{ij}$

For now we restrict the discussion to the case where $\boldsymbol{X}_t$ is non-random and thus we write $\boldsymbol{x}_t$ instead of $\boldsymbol{X}_t$.

# Outline of the lecture

- Regression based methods, 1st part:
  - **Ordinary Least Squares (OLS)**
  -
  -
  -

# Ordinary Least Squares (OLS)

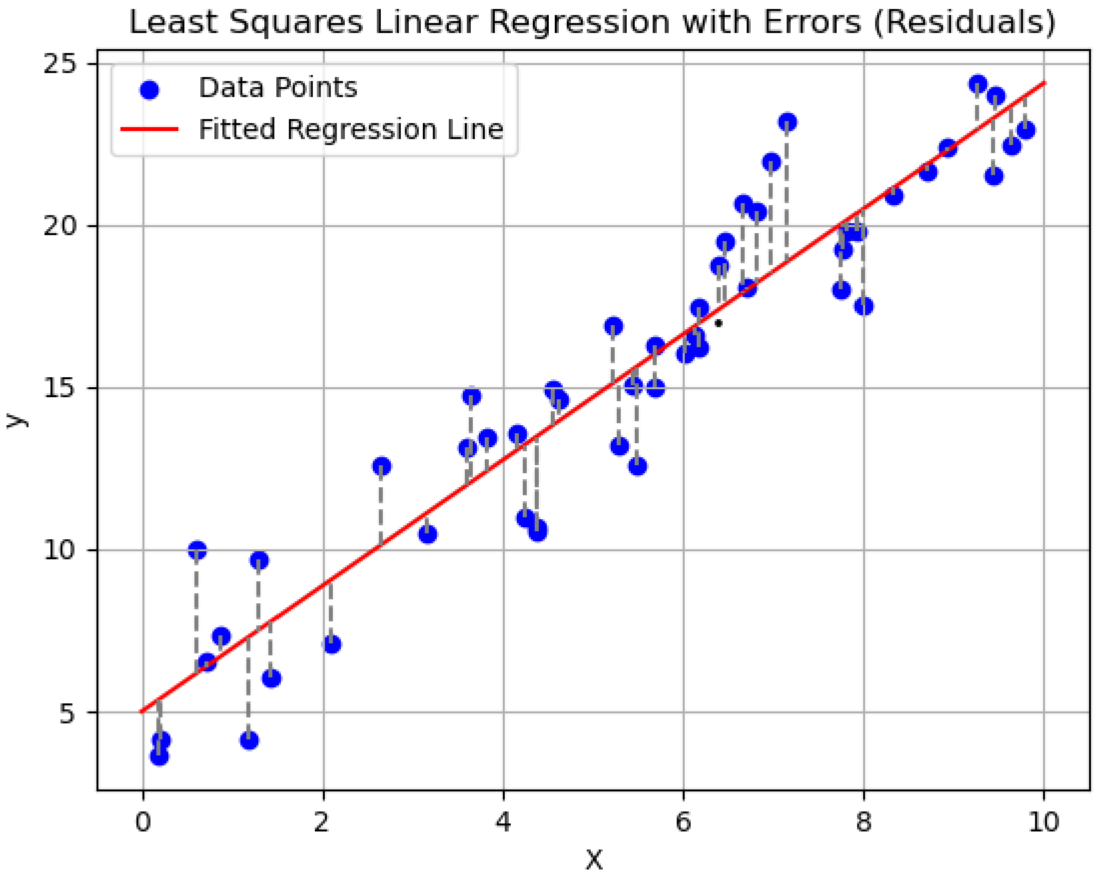
Observations (data):

$$(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \cdots, (y_n, \boldsymbol{x}_n)$$

OLS estimates (of the parameters) are found by minimizing the sum of squared residuals:

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta})$$

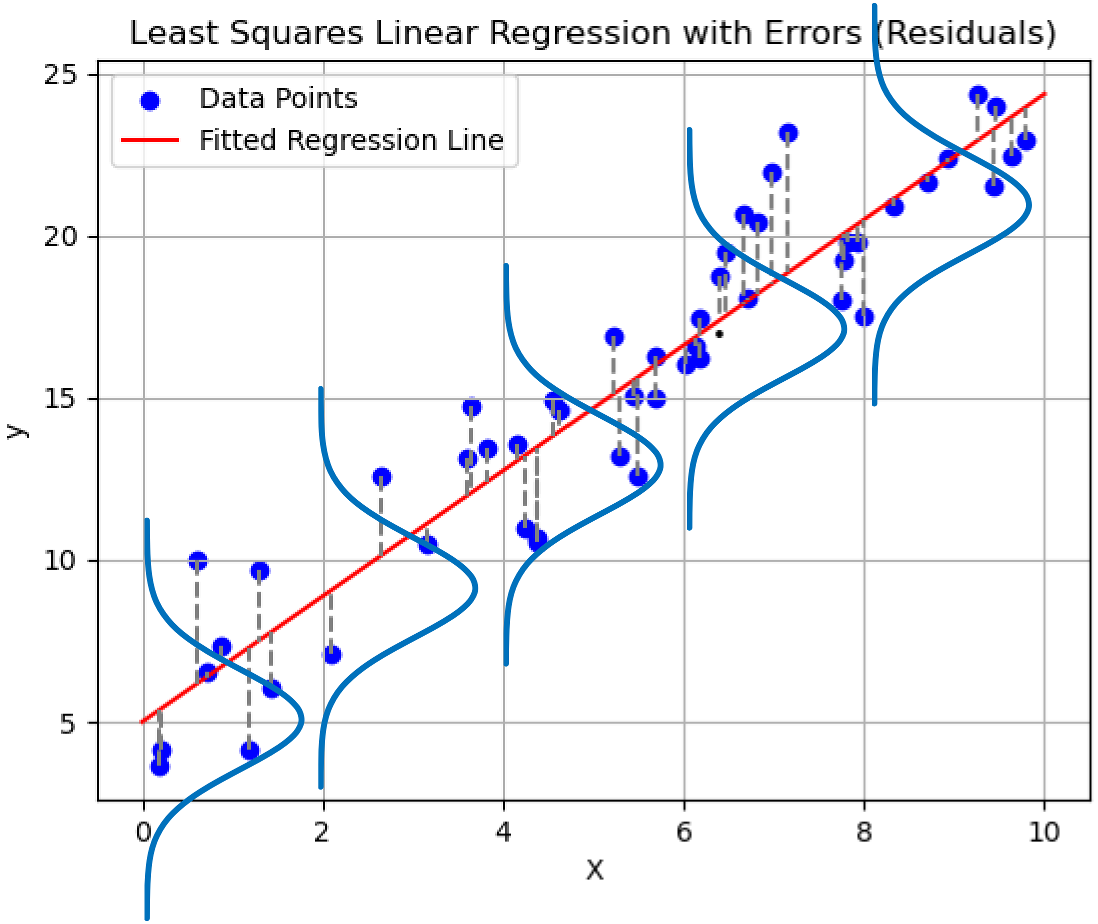
(sometimes also called ”RSS”)



# OLS - model assumptions

Errors must be assumed to all have the same variance and be mutually uncorrelated

Errors are "i.i.d"



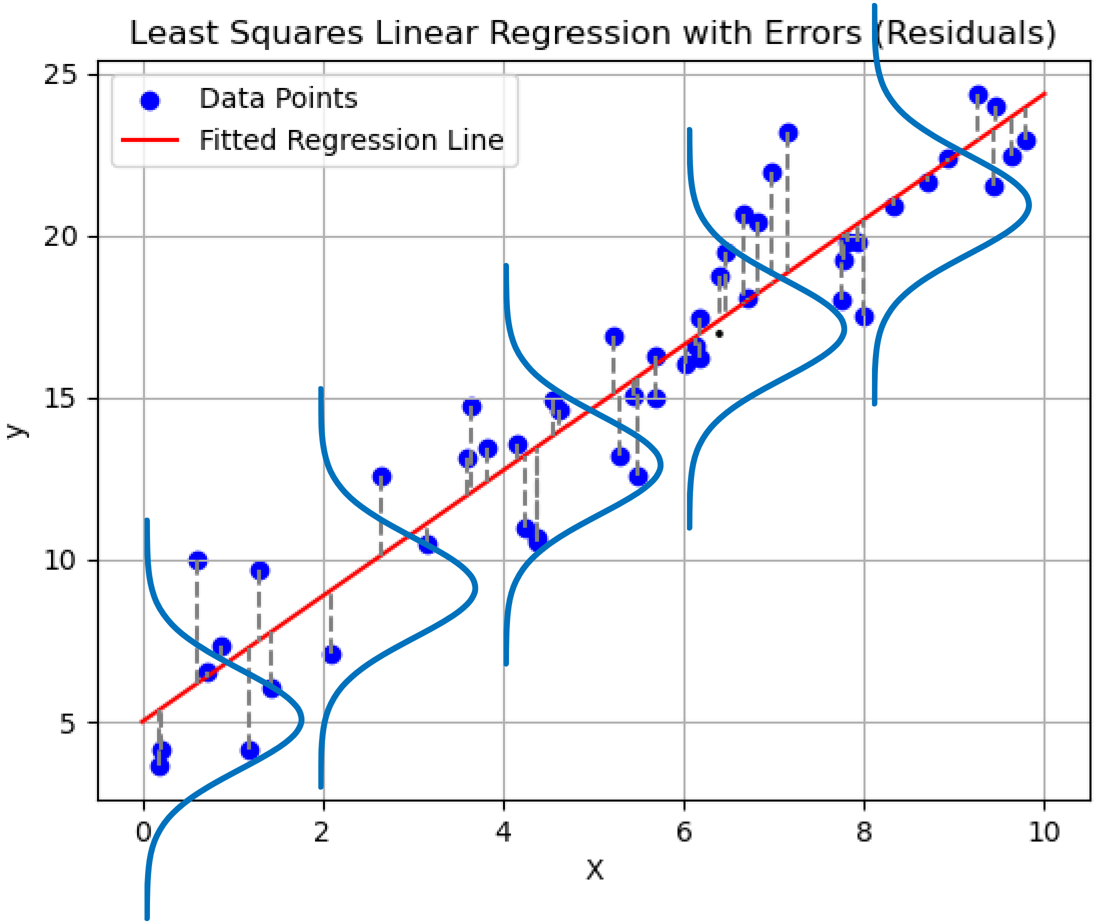
# OLS - model assumptions

Errors must be assumed to all have the same variance and be mutually uncorrelated

Errors are ”i.i.d”

$$\mathrm{Cov}[\varepsilon_{t_i}, \varepsilon_{t_j}] = \sigma_t^2 \Sigma_{ij}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$



# OLS – variance of model errors and of parameters

The variance of the model errors is estimated as:

$$\widehat{\sigma}^2 = \frac{S(\widehat{\boldsymbol{\theta}})}{n-p}$$

where $p$ is the number of estimated parameters.

The variance-covariance matrix of the estimates is approximately

$$V[\widehat{\boldsymbol{\theta}}] = 2\widehat{\sigma}^2 \left[ \frac{\partial^2}{\partial^2 \boldsymbol{\theta}} S(\boldsymbol{\theta}) \right]^{-1} \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

# The General Linear Model (GLM)

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

# The General Linear Model (GLM)

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Is this model a GLM?

$$Y_t = \theta_0 + \theta_1 z_t + \varepsilon_t$$

# The General Linear Model (GLM)

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Is this model a GLM?

$$Y_t = \theta_0 + \theta_1 z_t + \varepsilon_t$$

Yes, what about this one?

$$Y_t = \theta_0 + \theta_1 z_t + \theta_2 z_t^2 + \varepsilon_t$$

# The General Linear Model (GLM)

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Is this model a GLM?

$$Y_t = \theta_0 + \theta_1 z_t + \varepsilon_t$$

Yes, what about this one?

$$Y_t = \theta_0 + \theta_1 z_t + \theta_2 z_t^2 + \varepsilon_t$$

Also yes, since it can be written as

$$y_t = \begin{pmatrix} 1 & z_t & z_t^2 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} + \varepsilon_t$$

# The General Linear Model (GLM)

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Is this model a GLM?

$$Y_t = \theta_0 + \theta_1 z_t + \varepsilon_t$$

Yes, what about this one?

$$Y_t = \theta_0 + \theta_1 z_t + \theta_2 z_t^2 + \varepsilon_t$$

Also yes, since it can be written as

$$y_t = \begin{pmatrix} 1 & z_t & z_t^2 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} + \varepsilon_t$$

It is linearity in the parameters that matters!

# Matrix notation

For all observations the model equations are written as:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad or \quad \boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

”Design Matrix”

# Matrix notation

For all observations the model equations are written as:

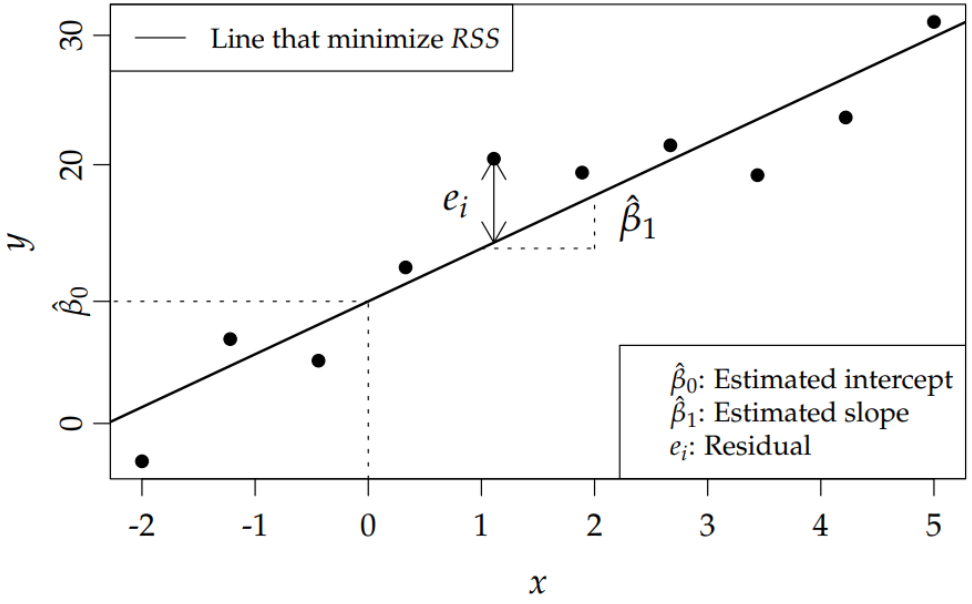
$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad or \quad \boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

”Design Matrix”

Ex: Simple linear regression:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = \{1, \ldots, n\}$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$



# OLS estimates for the general linear model

We minimise $S(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{x\theta})^T(\boldsymbol{Y} - \boldsymbol{x\theta})$, by solving $\frac{\partial}{\partial\widehat{\boldsymbol{\theta}}}S(\widehat{\boldsymbol{\theta}}) = 0$.

$$S(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{x\theta})^T(\boldsymbol{Y} - \boldsymbol{x\theta})$$

$$\nabla_\theta S(\boldsymbol{\theta}) = -2x^T(\boldsymbol{Y} - \boldsymbol{x\theta}) = \boldsymbol{0}$$

$$\boldsymbol{x}^T\boldsymbol{x\theta} = \boldsymbol{x}^T\boldsymbol{Y}$$

The solution is $\boxed{\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{Y}}$

# OLS estimates for the general linear model

Point estimate of parameters:

$$\boxed{\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{Y}}$$

Variance of parameters:

$$\mathrm{Var}[\widehat{\boldsymbol{\theta}}] = \mathrm{E}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] = \sigma^2 (\boldsymbol{x}^T \boldsymbol{x})^{-1}$$

Or approximately:

$$V[\widehat{\boldsymbol{\theta}}] = 2\widehat{\sigma}^2 \left[\frac{\partial^2}{\partial^2 \boldsymbol{\theta}} S(\boldsymbol{\theta})\right]^{-1}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

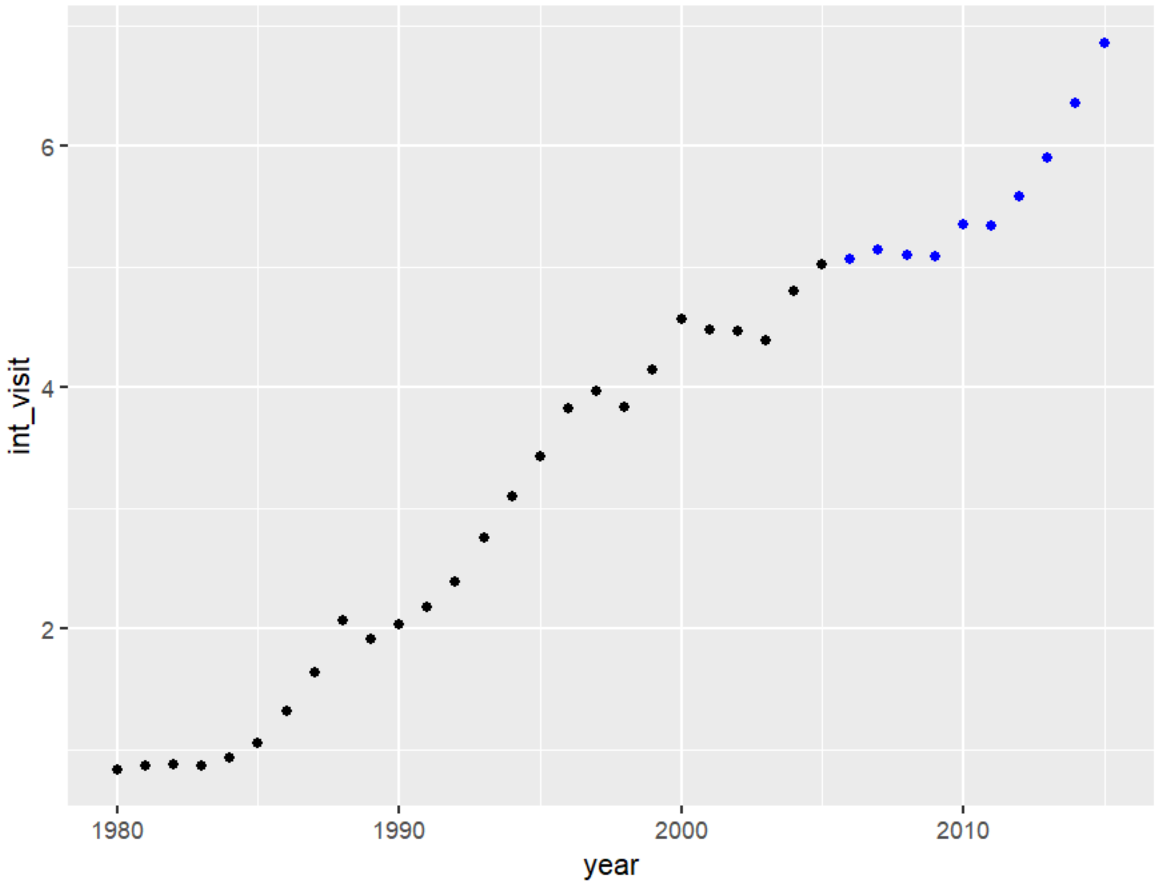
Estimate of the variance of residuals:

$$\widehat{\sigma}^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}/(n - p)$$

# Properties of the OLS-estimator of a GLM

- It is a linear function of the observations $\boldsymbol{Y}$ (and $\widehat{\boldsymbol{Y}}$ is thus a linear function of the observations)
- It is unbiased, i.e. $E[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$
- $V[\widehat{\boldsymbol{\theta}}] = E\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] = \sigma^2(\boldsymbol{x}^T\boldsymbol{x})^{-1}$
- $\hat{\theta}$ is BLUE (Best Linear Unbiased Estimator), which means that it has the smallest variance among all estimators which are a linear function of the observations.

# Example in R



# Example in R

```
> print(X)
      [,1] [,2]
 [1,]    1 1980
 [2,]    1 1981
 [3,]    1 1982
 [4,]    1 1983
 [5,]    1 1984
 [6,]    1 1985
 [7,]    1 1986
 [8,]    1 1987
 [9,]    1 1988
[10,]    1 1989
[11,]    1 1990
[12,]    1 1991
[13,]    1 1992
[14,]    1 1993
[15,]    1 1994
[16,]    1 1995
[17,]    1 1996
[18,]    1 1997
[19,]    1 1998
[20,]    1 1999
[21,]    1 2000
[22,]    1 2001
[23,]    1 2002
[24,]    1 2003
[25,]    1 2004
[26,]    1 2005
```

```
> print(y)
           [,1]
 [1,] 0.8298943
 [2,] 0.8595109
 [3,] 0.8766892
 [4,] 0.8667072
 [5,] 0.9320520
 [6,] 1.0482636
 [7,] 1.3111932
 [8,] 1.6375623
 [9,] 2.0641074
[10,] 1.9126828
[11,] 2.0354457
[12,] 2.1772113
[13,] 2.3896834
[14,] 2.7505921
[15,] 3.0906664
[16,] 3.4266403
[17,] 3.8306491
[18,] 3.9719086
[19,] 3.8316004
[20,] 4.1431010
[21,] 4.5665510
[22,] 4.4754100
[23,] 4.4627960
[24,] 4.3848290
[25,] 4.7968610
[26,] 5.0150490
```

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{Y}$$

```
OLS <- solve(t(X)%*%X)%*%t(X)%*%y
```

```
theta_0 <- OLS[1]
theta_1 <- OLS[2]
```

Parameter estimates

# Example in R

```
> print(X)
      [,1] [,2]
 [1,]    1 1980
 [2,]    1 1981
 [3,]    1 1982
 [4,]    1 1983
 [5,]    1 1984
 [6,]    1 1985
 [7,]    1 1986
 [8,]    1 1987
 [9,]    1 1988
[10,]    1 1989
[11,]    1 1990
[12,]    1 1991
[13,]    1 1992
[14,]    1 1993
[15,]    1 1994
[16,]    1 1995
[17,]    1 1996
[18,]    1 1997
[19,]    1 1998
[20,]    1 1999
[21,]    1 2000
[22,]    1 2001
[23,]    1 2002
[24,]    1 2003
[25,]    1 2004
[26,]    1 2005
```

```
> print(y)
          [,1]
 [1,] 0.8298943
 [2,] 0.8595109
 [3,] 0.8766892
 [4,] 0.8667072
 [5,] 0.9320520
 [6,] 1.0482636
 [7,] 1.3111932
 [8,] 1.6375623
 [9,] 2.0641074
[10,] 1.9126828
[11,] 2.0354457
[12,] 2.1772113
[13,] 2.3896834
[14,] 2.7505921
[15,] 3.0906664
[16,] 3.4266403
[17,] 3.8306491
[18,] 3.9719086
[19,] 3.8316004
[20,] 4.1431010
[21,] 4.5665510
[22,] 4.4754100
[23,] 4.4627960
[24,] 4.3848290
[25,] 4.7968610
[26,] 5.0150490
```

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{Y}$$

```
OLS <- solve(t(X)%*%X)%*%t(X)%*%y
```

```
theta_0 <- OLS[1]
theta_1 <- OLS[2]
```

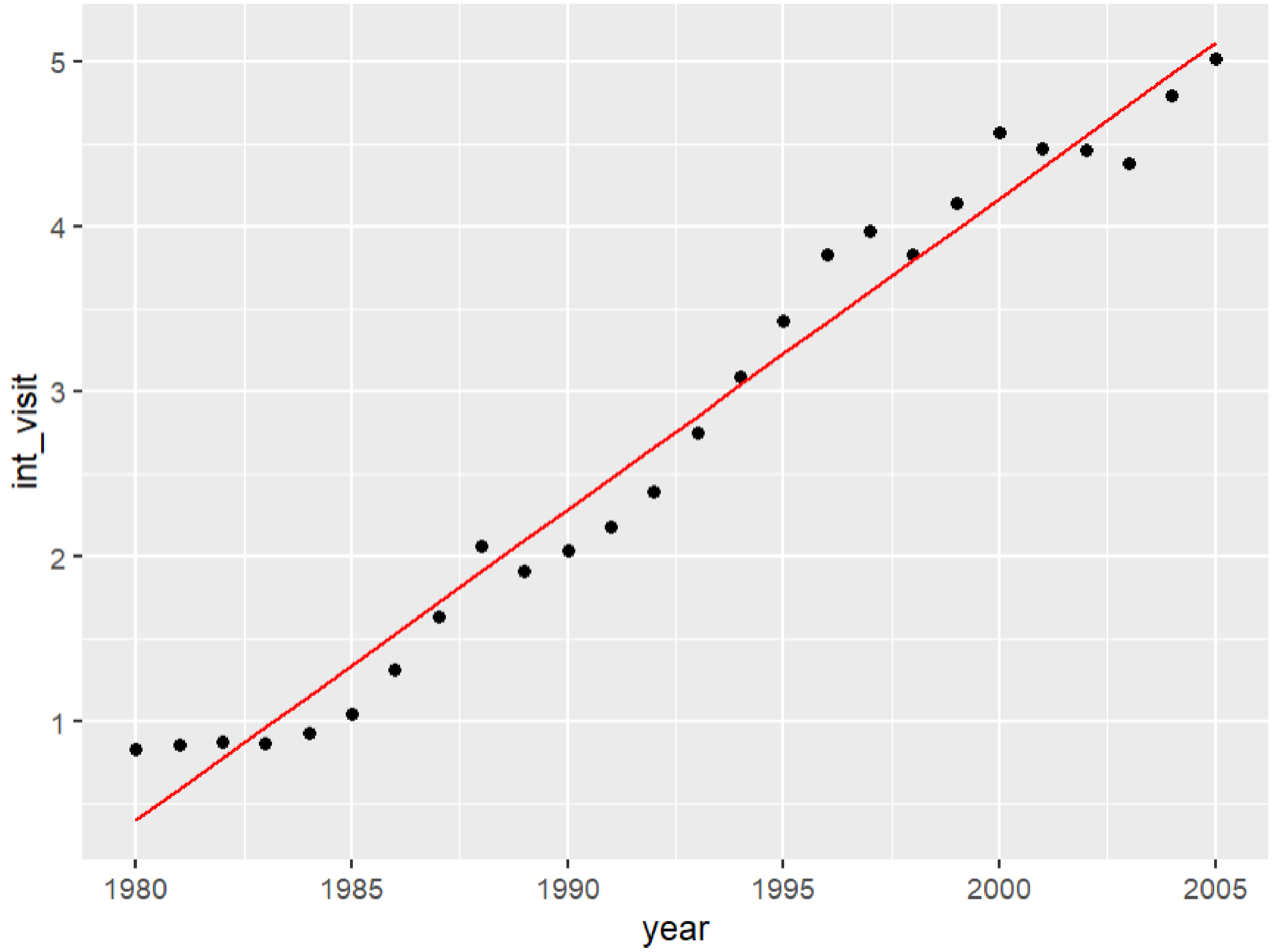
Parameter estimates

$$\hat{Y} = x\hat{\theta}$$

```
yhat_ols <- X%*%OLS
```

Predicted values

# Example in R



# Example in R

```
e_ols <- y - yhat_ols
```

```
RSS_ols <- t(e_ols)%*%e_ols
```

```
sigma2_ols <- as.numeric(RSS_ols/(n - nparams))
```

$$\widehat{\sigma}^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}/(n-p)$$

```
V_ols <- sigma2_ols * solve(t(X) %*% X)
```

$$\mathrm{Var}[\widehat{\boldsymbol{\theta}}] = \mathrm{E}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] = \sigma^2(\boldsymbol{x}^T\boldsymbol{x})^{-1}$$

```
> print(V_ols)
             [,1]          [,2]
[1,] 165.6609970 -8.314110e-02
[2,]  -0.0831411  4.172703e-05
```

```
se_theta_0 <- (sqrt(diag(V_ols)))[1]
se_theta_1 <- (sqrt(diag(V_ols)))[2]
```

Now we have estimated the parameters AND their standard error

# Outline of the lecture

- Regression based methods, 1st part:
  - Ordinary Least Squares (OLS)
  - **Predictions = forecast**
  -
  -

# Prediction in the general linear model

If the expected value of the squared prediction error is to be minimized, then the expected mean $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is the optimal predictor.

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

# Prediction in the general linear model

If the expected value of the squared prediction error is to be minimized, then the expected mean $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is the optimal predictor.

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Known parameters:

$$\widehat{Y}_t = E_\theta[Y_t | \boldsymbol{X}_t = \boldsymbol{x}_t] = \boldsymbol{x}_t^T \boldsymbol{\theta}$$

$$V_\theta[Y_t - \widehat{Y}_t] = V_\theta[\varepsilon_t] = \sigma^2$$

# Prediction in the general linear model

If the expected value of the squared prediction error is to be minimized, then the expected mean $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is the optimal predictor.

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \varepsilon_t$$

Known parameters:

$$\widehat{Y}_t = E_\theta[Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t] = \boldsymbol{x}_t^T \boldsymbol{\theta}$$

$$V_\theta[Y_t - \widehat{Y}_t] = V_\theta[\varepsilon_t] = \sigma^2$$

Estimated parameters:

$$\widehat{Y}_t = E_{\widehat{\theta}}[Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t] = \boldsymbol{x}_t^T \widehat{\boldsymbol{\theta}}$$

$$V_{\widehat{\theta}}[Y_t - \widehat{Y}_t] = V_{\widehat{\theta}}[\varepsilon_t + \boldsymbol{x}_t^T(\theta - \widehat{\theta})] = \widehat{\sigma}^2[1 + \boldsymbol{x}_t^T(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}_t]$$

# Prediction in the general linear model, continued

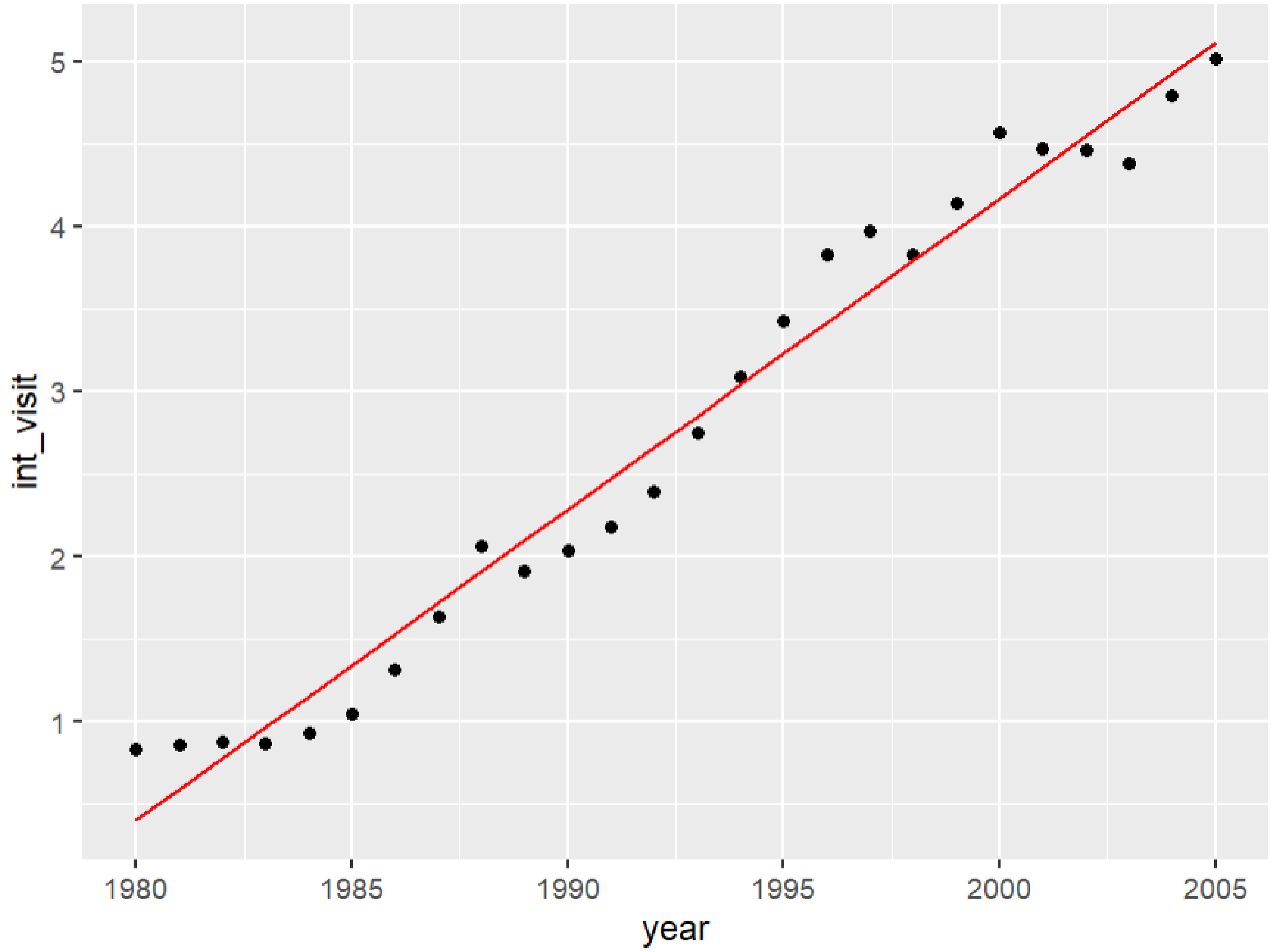
Prediction interval for the predicted values:

$$\widehat{Y}_t \pm t_{\alpha/2}(n-p)\hat{\sigma}\sqrt{1 + \boldsymbol{x}_t^T(\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}_t}$$

Here $t_{\alpha/2}(n-p)$ refers to a percentile in the t-distribution with n-p degrees of freedom.
If n-p is large, percentiles from the normal distribution can be used.

In time series analysis: Prediction of future values = **”forecast”**

# Example in R




Lets make predictions (forecast) for the next 10 years

# Example in R

```
> print(Xtest)
      [,1] [,2]
 [1,]    1 2006
 [2,]    1 2007
 [3,]    1 2008
 [4,]    1 2009
 [5,]    1 2010
 [6,]    1 2011
 [7,]    1 2012
 [8,]    1 2013
 [9,]    1 2014
[10,]    1 2015
```

$$\widehat{Y}_t = E_{\widehat{\theta}}[Y_t | \boldsymbol{X}_t = \boldsymbol{x}_t] = \boldsymbol{x}_t^T \widehat{\boldsymbol{\theta}}$$

```
y_pred <- Xtest%*%OLS
```

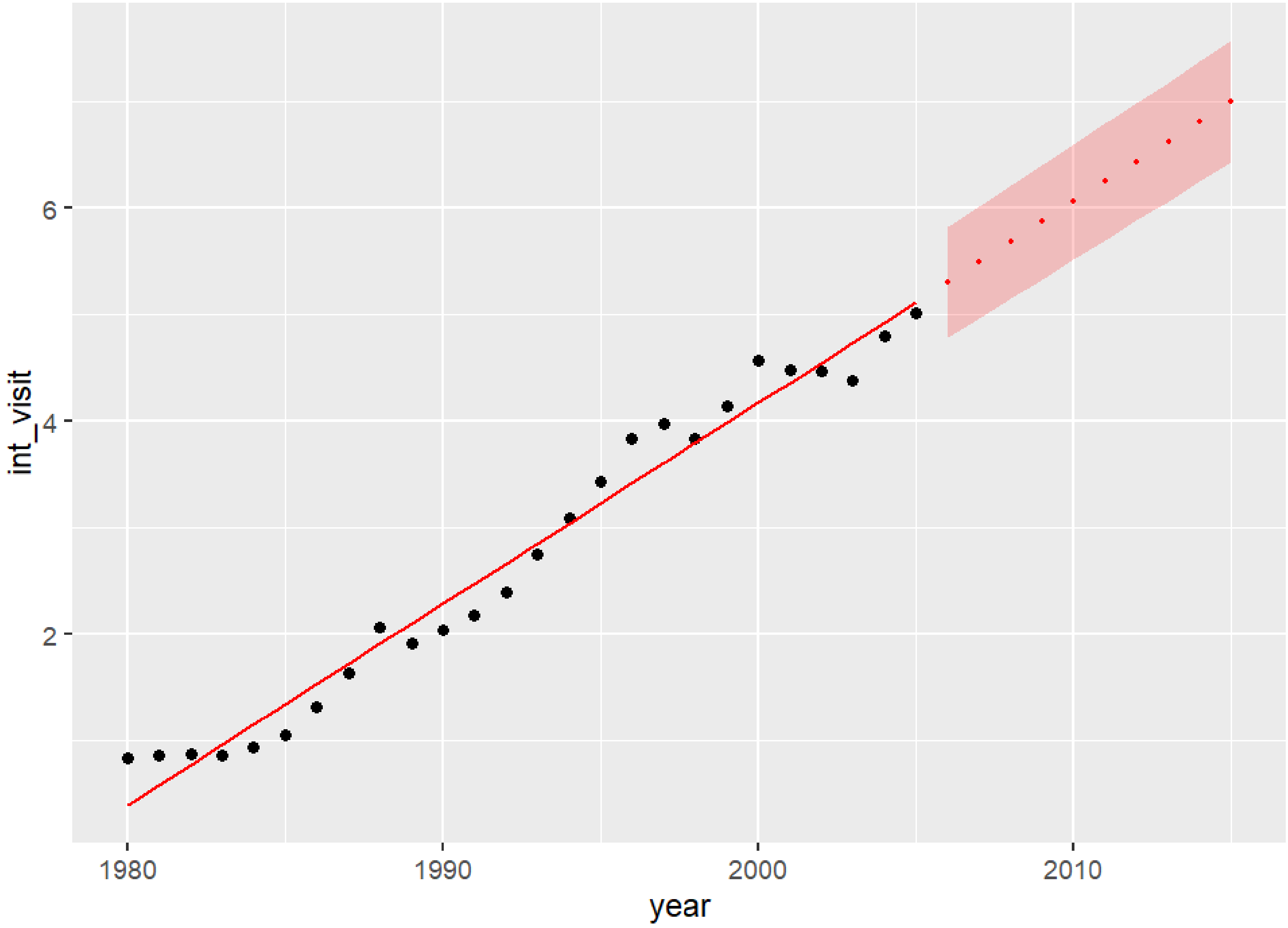
$$\widehat{\sigma}^2[1 + \boldsymbol{x}_t^T (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}_t]$$

```
Vmatrix_pred <- sigma2_ols*(1+(Xtest%*%solve(t(X)%*%X))%*%t(Xtest))
```

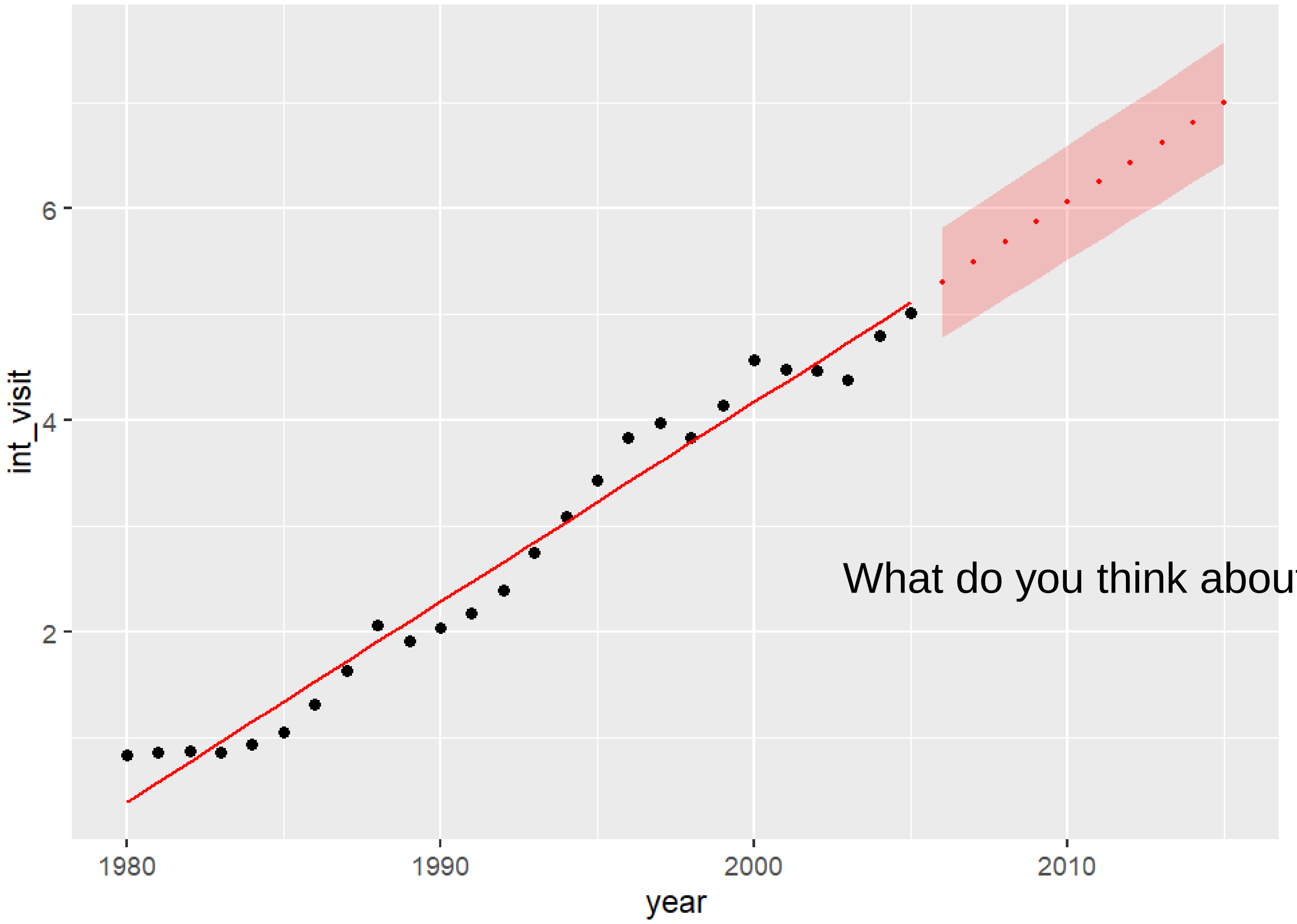
$$\widehat{Y}_t \pm t_{\alpha/2}(n-p)\hat{\sigma}\sqrt{1 + \boldsymbol{x}_t^T (\boldsymbol{x}^T \boldsymbol{x})^{-1} \boldsymbol{x}_t}$$

```
y_pred_lwr <- y_pred - 1.96*sqrt(diag(Vmatrix_pred))
y_pred_upr <- y_pred + 1.96*sqrt(diag(Vmatrix_pred))
```

# Example in R

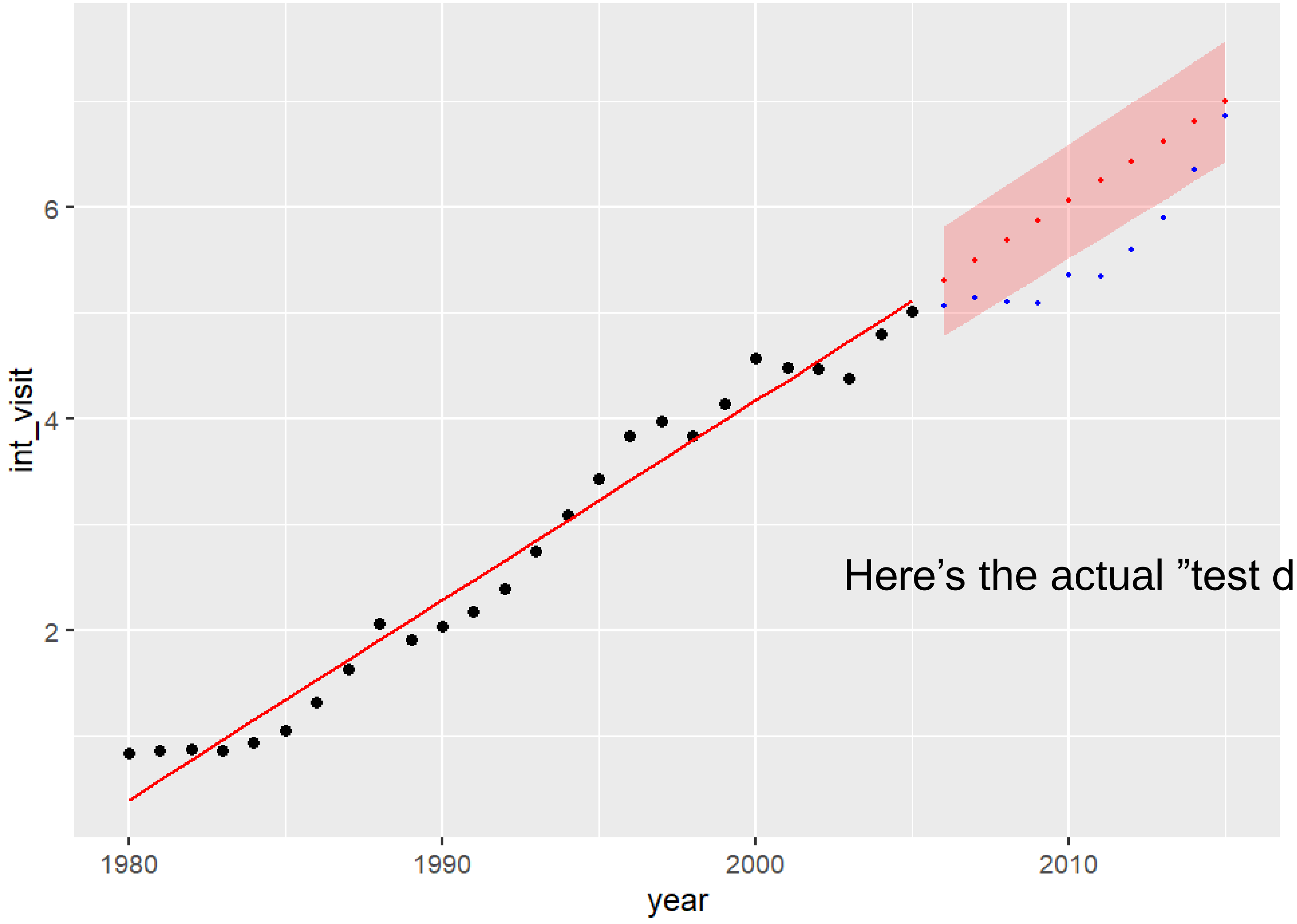


# Example in R




What do you think about the forecast?

# Example in R




Here's the actual "test data"

# Outline of the lecture

- Regression based methods, 1st part:
  - Ordinary Least Squares (OLS)
  - Predictions = forecast
  - **Global Trend Model**
  -

# Sneakpeak at “Trend Models”

Trend models are:

- Linear regression models
- The independent variables are functions of time
- The reference time is often the latest timepoint instead of the ”origin”
- Notation is a bit different – the principle is the same

$$Y_{N+j} = f^T(j)\boldsymbol{\theta} + \varepsilon_{N+j}$$

Today we will only talk about **global** trend models

# Sneakpeak at “Trend Models”

Trend models are:

- Linear regression models
- The independent variables are functions of time
- The reference time is often the latest timepoint instead of the ”origin”
- Notation is a bit different – the principle is the same

$$Y_{N+j} = f^T(j)\boldsymbol{\theta} + \varepsilon_{N+j}$$

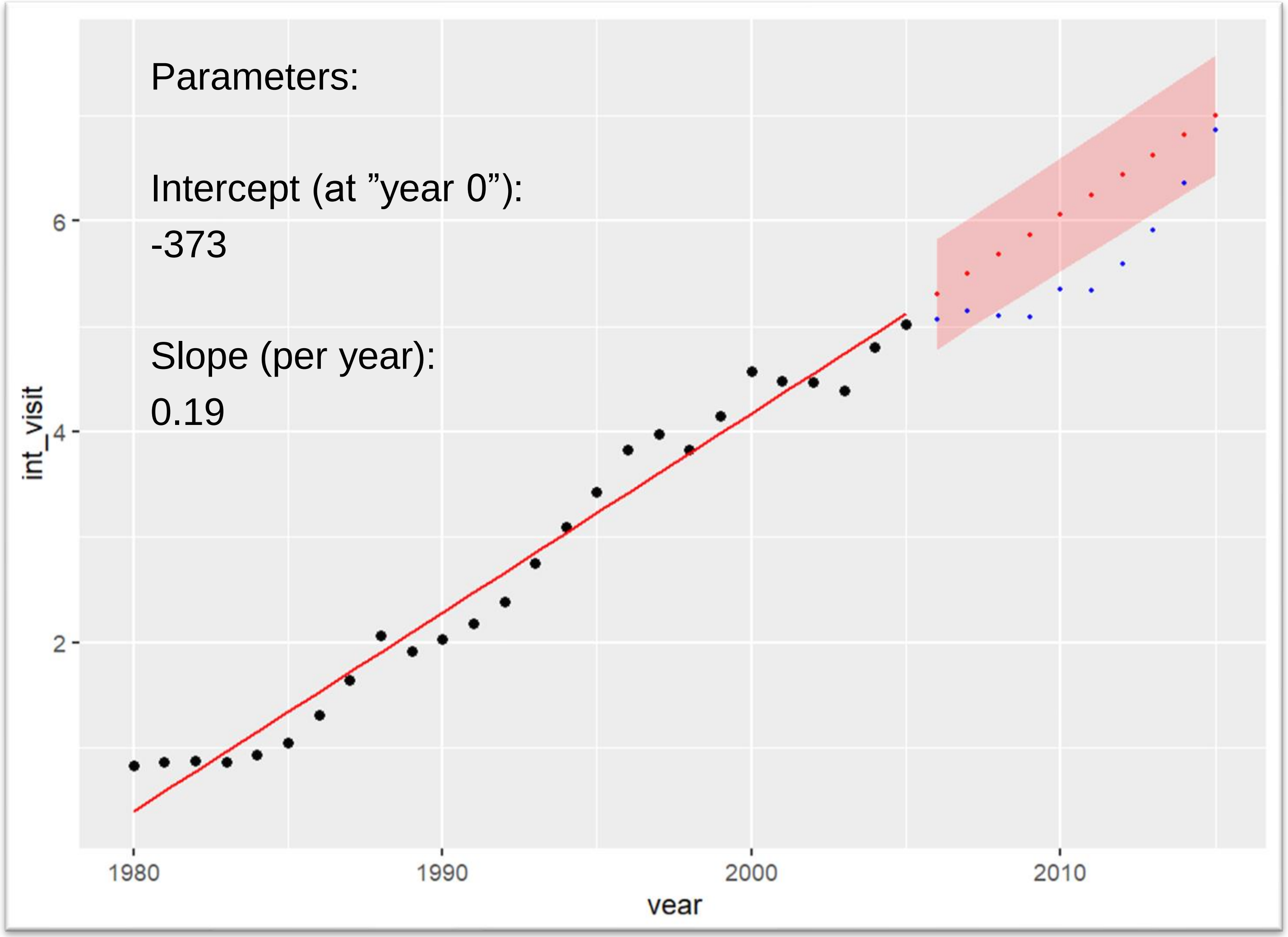
N refers to the ”latest” timepoint in the data – this is our reference timepoint (=”now”)

If we put j=0, we get the estimate ”now”

j = 1,2,3,… refers to future timepoints

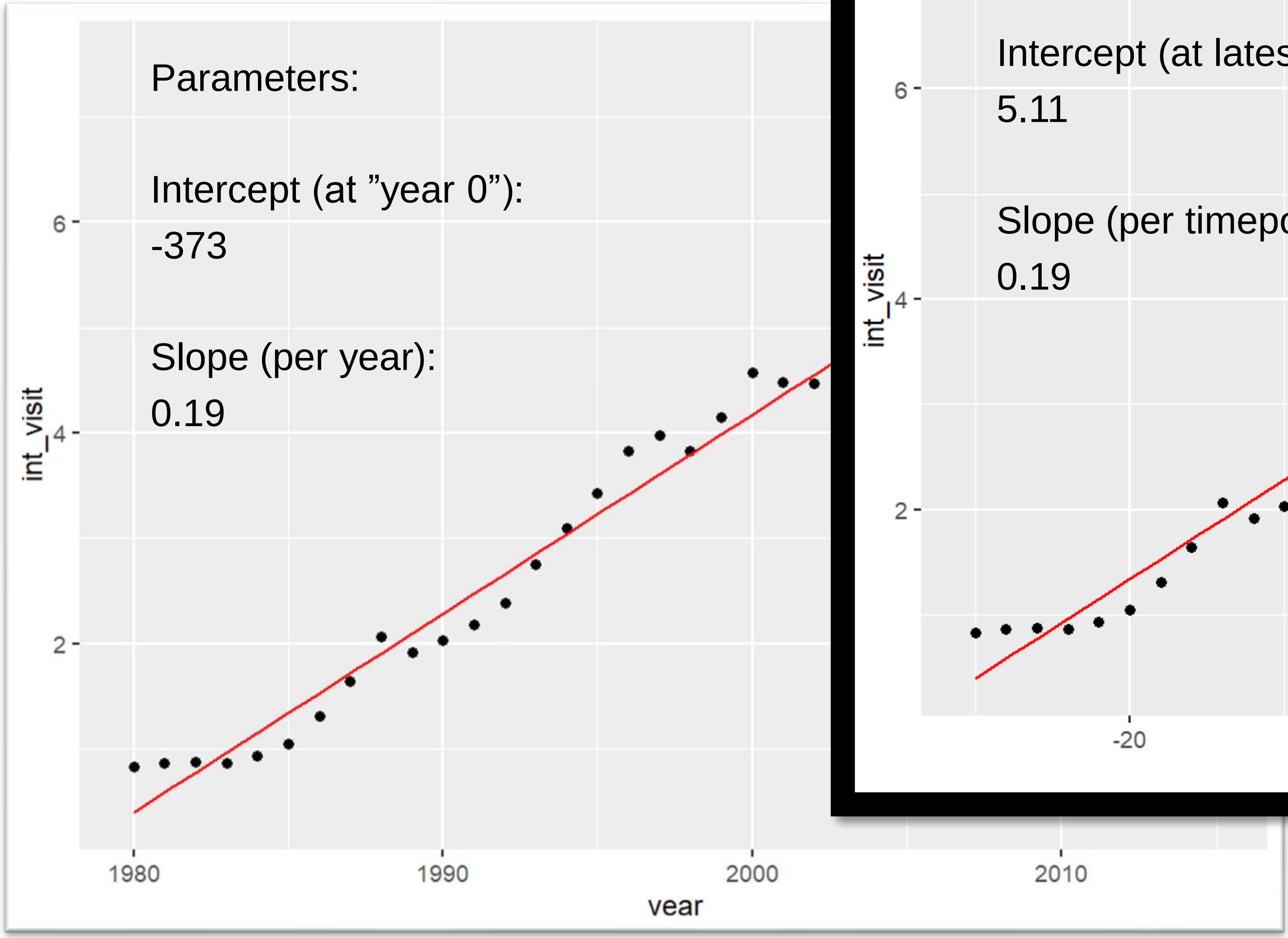
The data available to estimate the model is the data from j = 0, -1, -2, -3, …

# Example in R

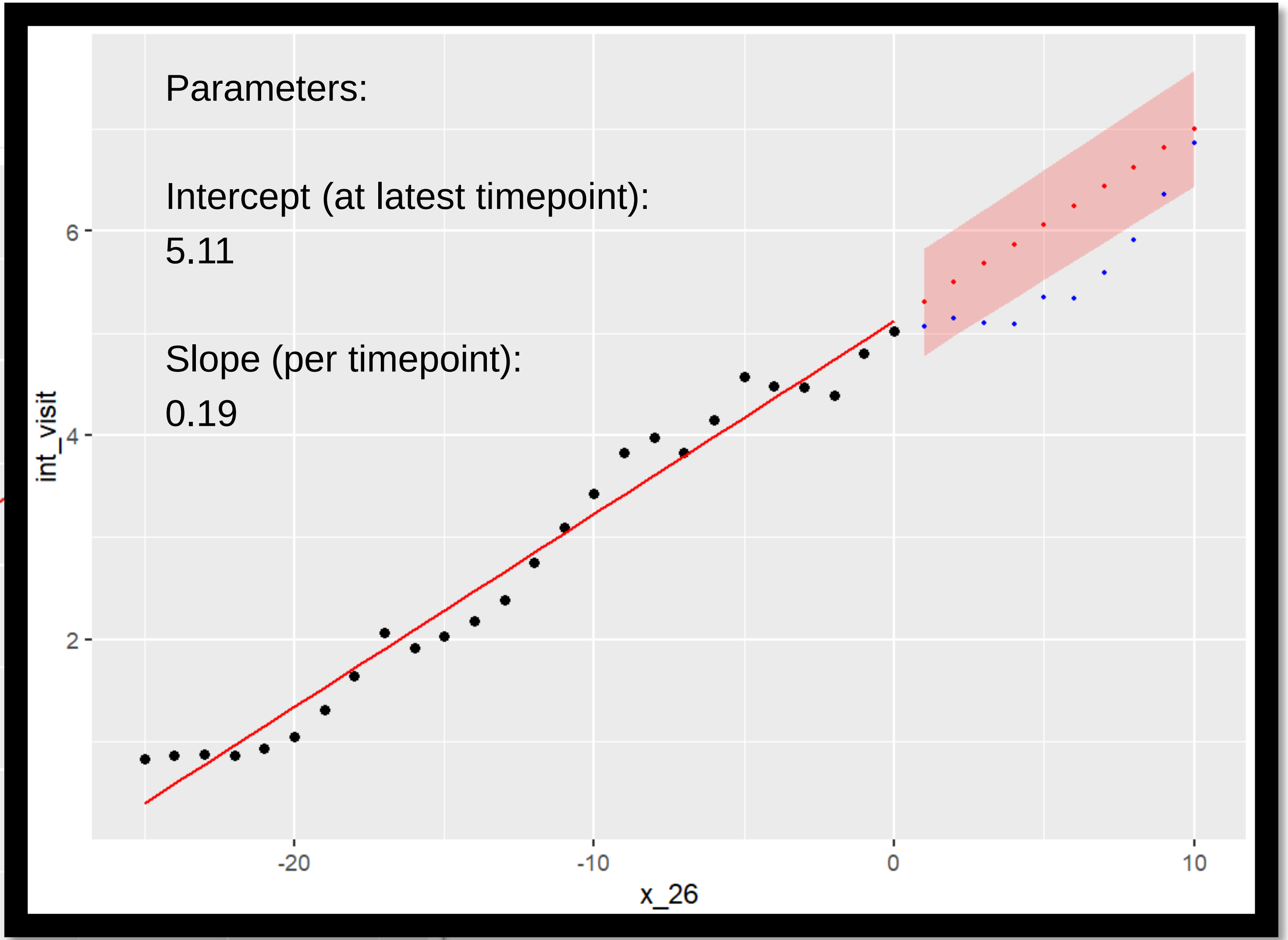


Parameters:
Intercept (at ”year 0”):
-373
Slope (per year):
0.19
int_visit
6
4
2
1980
1990
2000
2010
year

# Example in R




Trend model setting:




Notice the change in x-axis!

Number of obs. in training data = N = 26c

# The linear trend model

- The general trend model:

$$Y_{N+j} = \boldsymbol{f}^T(j)\boldsymbol{\theta} + \varepsilon_{N+j}$$

- The linear trend model is obtained when: $\boldsymbol{f}(j) = \begin{pmatrix} 1 \\ j \end{pmatrix}$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(-N+1) \\ \boldsymbol{f}^T(-N+2) \\ \vdots \\ \boldsymbol{f}^T(0) \end{bmatrix} \boldsymbol{\theta}_N + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

# Example in R

```
> print(X)
      [,1] [,2]
 [1,]    1 1980
 [2,]    1 1981
 [3,]    1 1982
 [4,]    1 1983
 [5,]    1 1984
 [6,]    1 1985
 [7,]    1 1986
 [8,]    1 1987
 [9,]    1 1988
[10,]    1 1989
[11,]    1 1990
[12,]    1 1991
[13,]    1 1992
[14,]    1 1993
[15,]    1 1994
[16,]    1 1995
[17,]    1 1996
[18,]    1 1997
[19,]    1 1998
[20,]    1 1999
[21,]    1 2000
[22,]    1 2001
[23,]    1 2002
[24,]    1 2003
[25,]    1 2004
[26,]    1 2005
```

”Design-Matrix” changes →

```
> print(X_N)
      [,1] [,2]
 [1,]    1  -25
 [2,]    1  -24
 [3,]    1  -23
 [4,]    1  -22
 [5,]    1  -21
 [6,]    1  -20
 [7,]    1  -19
 [8,]    1  -18
 [9,]    1  -17
[10,]    1  -16
[11,]    1  -15
[12,]    1  -14
[13,]    1  -13
[14,]    1  -12
[15,]    1  -11
[16,]    1  -10
[17,]    1   -9
[18,]    1   -8
[19,]    1   -7
[20,]    1   -6
[21,]    1   -5
[22,]    1   -4
[23,]    1   -3
[24,]    1   -2
[25,]    1   -1
[26,]    1    0
```

Vector of y-values stays the same

```
> print(y)
           [,1]
 [1,] 0.8298943
 [2,] 0.8595109
 [3,] 0.8766892
 [4,] 0.8667072
 [5,] 0.9320520
 [6,] 1.0482636
 [7,] 1.3111932
 [8,] 1.6375623
 [9,] 2.0641074
[10,] 1.9126828
[11,] 2.0354457
[12,] 2.1772113
[13,] 2.3896834
[14,] 2.7505921
[15,] 3.0906664
[16,] 3.4266403
[17,] 3.8306491
[18,] 3.9719086
[19,] 3.8316004
[20,] 4.1431010
[21,] 4.5665510
[22,] 4.4754100
[23,] 4.4627960
[24,] 4.3848290
[25,] 4.7968610
[26,] 5.0150490
```

# Trend model, OLS estimates

$$\boldsymbol{Y}_N = \boldsymbol{x}_N \boldsymbol{\theta}_N + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(-N+1) \\ \boldsymbol{f}^T(-N+2) \\ \vdots \\ \boldsymbol{f}^T(0) \end{bmatrix} \boldsymbol{\theta}_N + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

$$\widehat{\boldsymbol{\theta}}_N = (\boldsymbol{x}_N^T \boldsymbol{x}_N)^{-1} \boldsymbol{x}_N^T \boldsymbol{Y}_N = \boldsymbol{F}_N^{-1} \boldsymbol{h}_N$$

$$\boldsymbol{F}_N = \boldsymbol{x}_N^T \boldsymbol{x}_N = \sum_{j=0}^{N-1} \boldsymbol{f}(-j) \boldsymbol{f}^T(-j)$$

$$\boldsymbol{h}_N = \boldsymbol{x}_N^T \boldsymbol{Y} = \sum_{j=0}^{N-1} \boldsymbol{f}(-j) Y_{N-j}$$

It is still an OLS regression model!

The notation is different

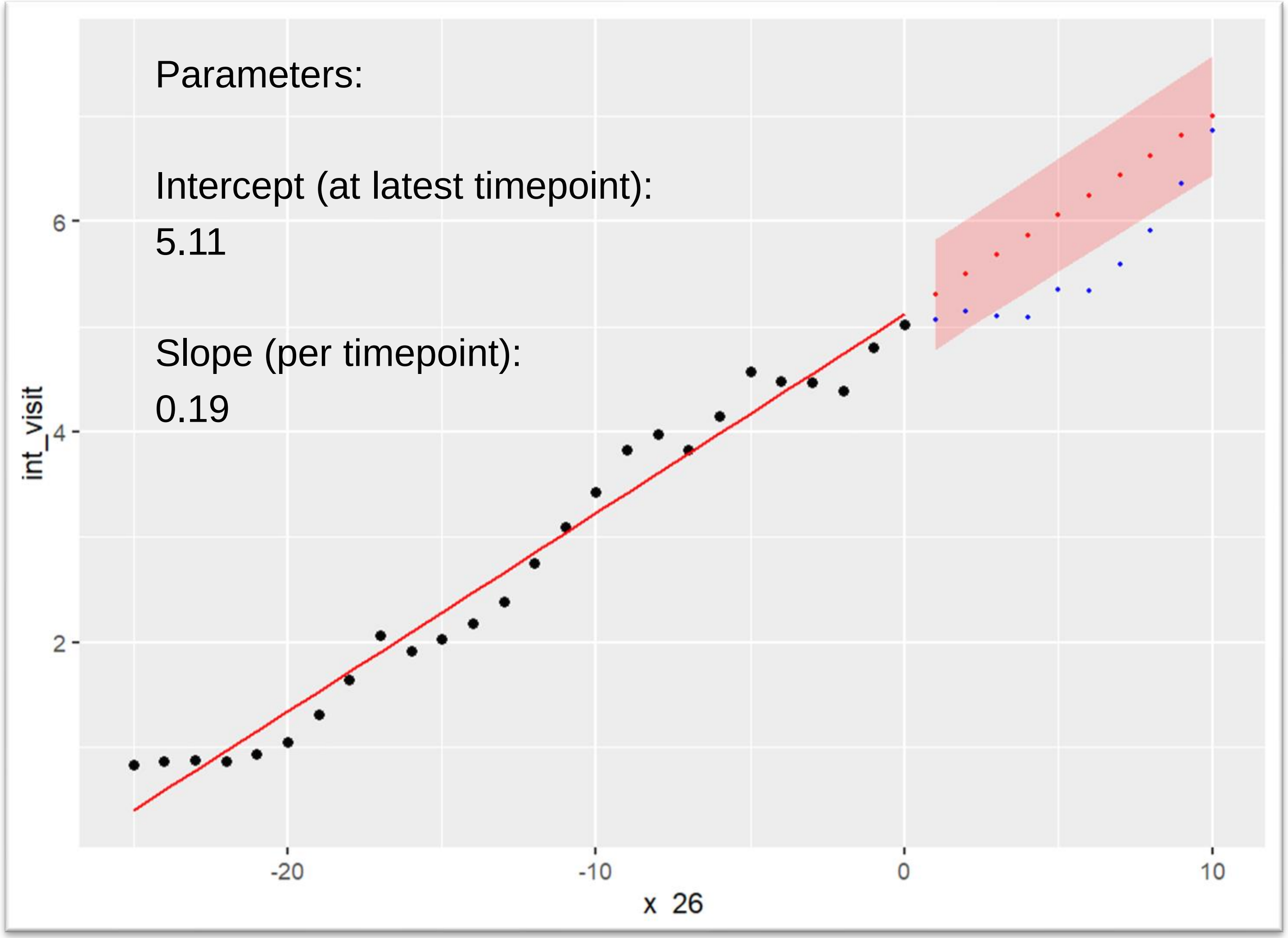
# Example in R

Estimating the parameters:

```
F_N <- t(X_N)%*%X_N
```

```
h_N <- t(X_N)%*%y
```

```
theta_N <- solve(F_N)%*%h_N
print(theta_N)
# [1,] 5.115330
# [2,] 0.188649
```



# Other trend models

- Constant mean: $Y_{N+j} = \theta_0 + \varepsilon_{N+j}$
- Linear trend: $Y_{N+j} = \theta_0 + \theta_1 j + \varepsilon_{N+j}$
- Quadratic trend: $Y_{N+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \varepsilon_{n+j}$
- $k$'th order polynomial trend: $Y_{n+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \cdots + \theta_k \frac{j^k}{k!} + \varepsilon_{N+j}$
- Harmonic model with the period $p$: $Y_{N+j} = \theta_0 + \theta_1 \sin \frac{2\pi}{p} j + \theta_2 \cos \frac{2\pi}{p} j + \varepsilon_{N+j}$

# Trend model forecasting ("$\ell$-step predictions")

- Prediction:

$$\widehat{Y}_{N+\ell|N} = \boldsymbol{f}^T(\ell)\widehat{\boldsymbol{\theta}}_N$$

- Variance of the prediction error:

$$V[Y_{N+\ell} - \widehat{Y}_{N+\ell|N}] = \sigma^2 \left[1 + \boldsymbol{f}^T(\ell)\boldsymbol{F}_N^{-1}\boldsymbol{f}(\ell)\right]$$

- $100(1-\alpha)\%$ prediction interval:

$$\begin{aligned} &\widehat{Y}_{N+\ell|N} \pm \mathrm{t}_{\alpha/2}(N-p)\sqrt{V[e_N(\ell)]} \\ =&\widehat{Y}_{N+\ell|N} \pm \mathrm{t}_{\alpha/2}(N-p)\widehat{\sigma}\sqrt{1 + \boldsymbol{f}^T(\ell)\boldsymbol{F}_N^{-1}\boldsymbol{f}(\ell)} \end{aligned}$$

where $\hat{\sigma}^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}/(N-p)$ ($p$ is the number of estimated parameters)

# Example in R

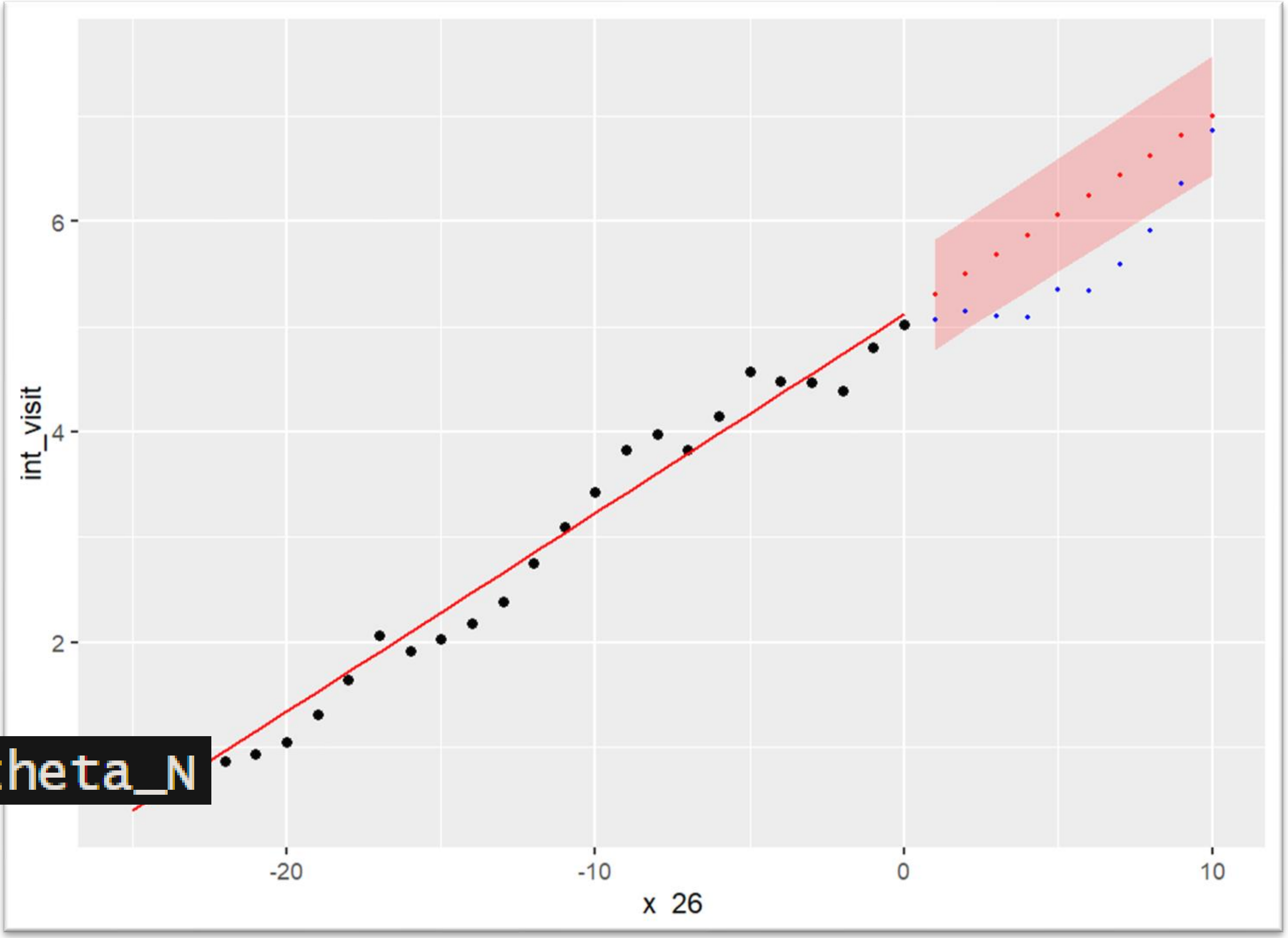
Predicting the future values
(here we predict for j = 1,2,3,4,5,6,7,8,9,10)

$$\boldsymbol{f}(j) = \begin{pmatrix} 1 \\ j \end{pmatrix}$$

```
f <- function(j) rbind(1, j)
```

$$\widehat{Y}_{N+\ell|N} = \boldsymbol{f}^T(\ell)\widehat{\boldsymbol{\theta}}_N$$

```
y_pred_N <- t(f(1:10)) %*% theta_N
```

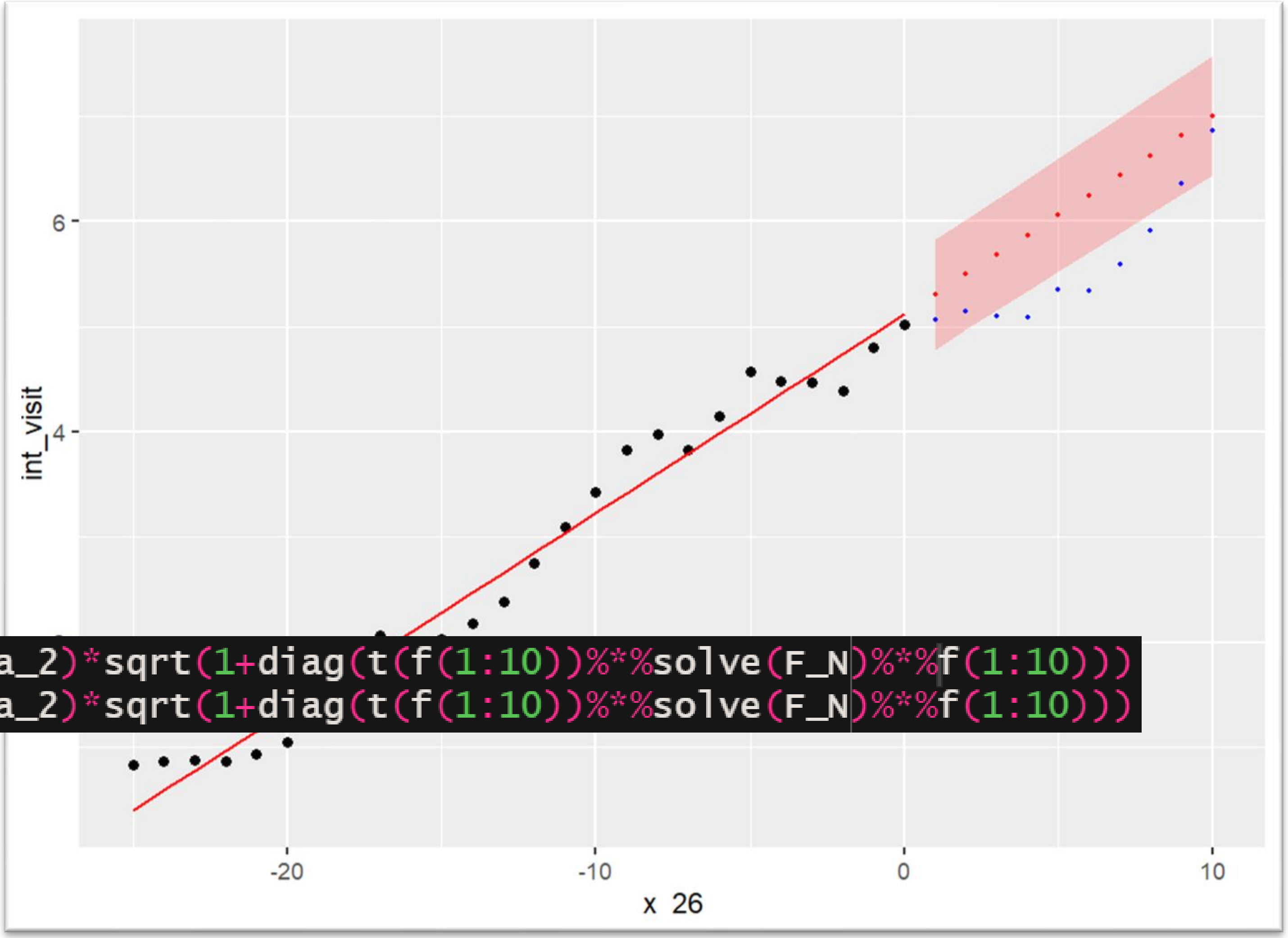


# Example in R

Compute prediction intervals

$$\widehat{Y}_{N+\ell|N} \pm t_{\alpha/2}(N-p)\widehat{\sigma}\sqrt{1+\boldsymbol{f}^T(\ell)\boldsymbol{F}_N^{-1}\boldsymbol{f}(\ell)}$$

```
e_N <- y - yhat_N
RSS <-  t(e_N)%*%e_N
sigma_2 <- as.numeric(RSS/(n-nparams))
```




```
y_pred_N_lwr <- y_pred_N - 1.96*sqrt(sigma_2)*sqrt(1+diag(t(f(1:10))%*%solve(F_N)%*%f(1:10)))
y_pred_N_upr <- y_pred_N + 1.96*sqrt(sigma_2)*sqrt(1+diag(t(f(1:10))%*%solve(F_N)%*%f(1:10)))
```

# Why do all this ??

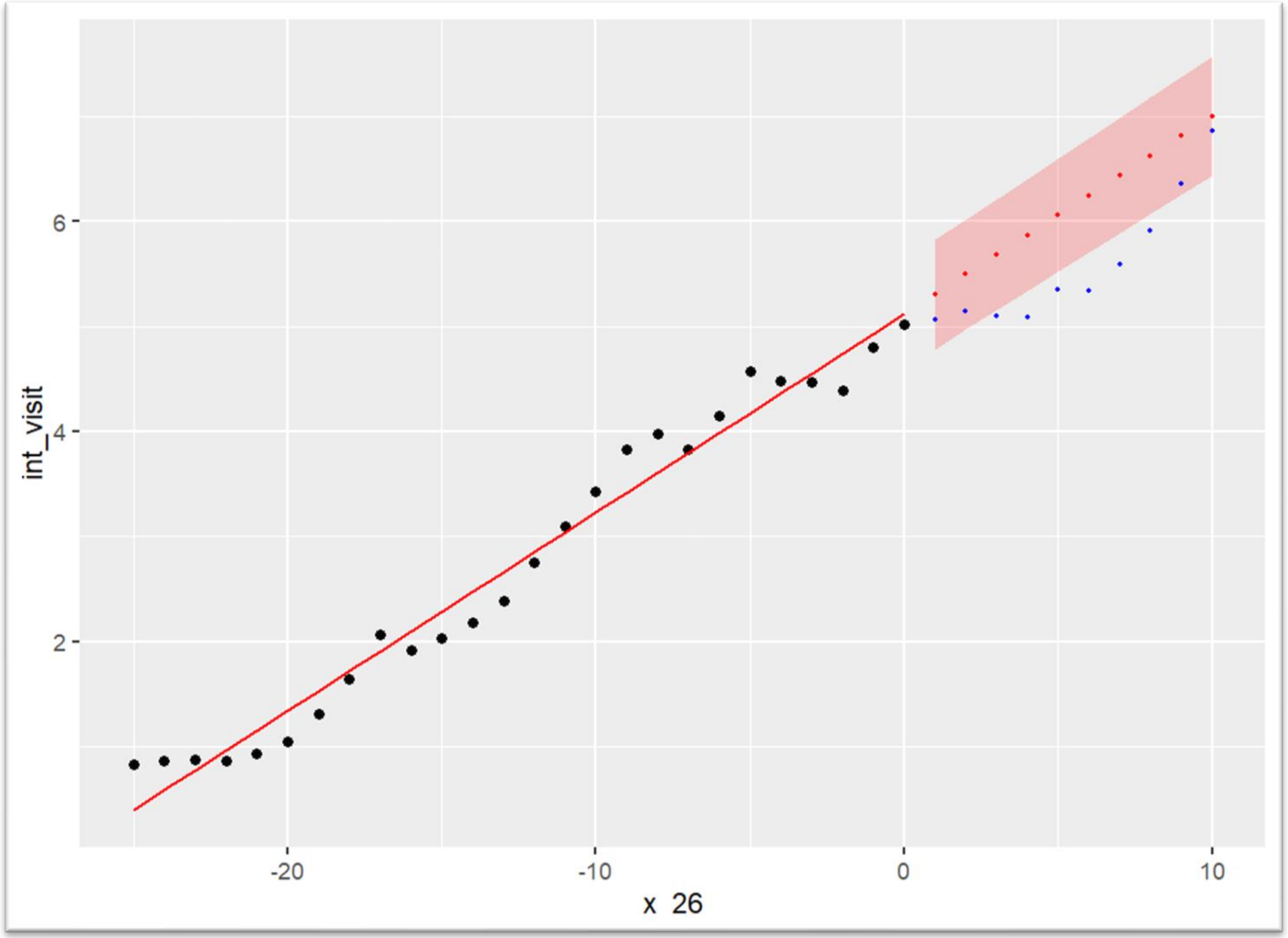
# Iterative updates

- Task:
  - Going from estimates based on $t = 1, \dots, N$, i.e. $\widehat{\boldsymbol{\theta}}_N$ to
  - estimates based on $t = 1, \dots, N, N+1$, i.e. $\widehat{\boldsymbol{\theta}}_{N+1}$
  - without redoing everything...
- Solution:

$$
\begin{aligned}
\boldsymbol{F}_{N+1} &= \boldsymbol{F}_N + \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \\
\boldsymbol{h}_{N+1} &= \boldsymbol{L}^{-1}\boldsymbol{h}_N + \boldsymbol{f}(0)Y_{N+1} \\
\widehat{\boldsymbol{\theta}}_{N+1} &= \boldsymbol{F}_{N+1}^{-1}\boldsymbol{h}_{N+1}
\end{aligned}
$$

Also iterative updates becomes smart, when we start doing **local** trend models (next week)

# Update the model,




We get the next observation

We want to shift the x-axis to the latest observation

And redo the regression line

And redo the forecast

# Updating the model

$$
\begin{aligned}
\boldsymbol{F}_{N+1} &= \boldsymbol{F}_N + \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \\
\boldsymbol{h}_{N+1} &= \boxed{\boldsymbol{L}^{-1}}\boldsymbol{h}_N + \boldsymbol{f}(0)\,Y_{N+1} \\
\widehat{\boldsymbol{\theta}}_{N+1} &= \boldsymbol{F}_{N+1}^{-1}\boldsymbol{h}_{N+1}
\end{aligned}
$$

$\boldsymbol{L}$ defines the transition from $\boldsymbol{f}(j)$ to $\boldsymbol{f}(j+1)$

$$\boldsymbol{f}(j+1) = \boldsymbol{L}\boldsymbol{f}(j)$$

# L matrix examples

## Linear trend model

$$Y_{N+j} = \theta_0 + \theta_1 j + \varepsilon_{N+j} \qquad \boldsymbol{L} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad \boldsymbol{f}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

## Quadratic trend model

$$Y_{N+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \varepsilon_{N+j} \qquad \boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1/2 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{f}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

# L matrix examples

**$k$'th order polynomial trend**

$$Y_{N+j} = \theta_0 + \theta_1 j + \theta_2 \frac{j^2}{2} + \cdots + \theta_k \frac{j^k}{k!} + \varepsilon_{N+j}$$

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ 1/2 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1/k! & 1/(k-1)! & & 1 \end{pmatrix}, \qquad \boldsymbol{f}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

# L matrix examples

**Harmonic model with the period $p$**

$$Y_{N+j} = \theta_0 + \theta_1 \sin\left(\frac{2\pi}{p}\right) j + \theta_2 \cos\left(\frac{2\pi}{p}\right) j + \varepsilon_{N+j}$$

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\left(\frac{2\pi}{p}\right) & \sin\left(\frac{2\pi}{p}\right) \\ 0 & -\sin\left(\frac{2\pi}{p}\right) & \cos\left(\frac{2\pi}{p}\right) \end{pmatrix}, \qquad \boldsymbol{f}(0) = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

# Example in R

```r
L <- matrix(c(1.,0., 1.,1.),
                  byrow=TRUE, nrow=2)
```

```
> print(L)
     [,1] [,2]
[1,]    1    0
[2,]    1    1
```

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

```r
F_N_27 <- F_N + f(-n) %*% t(f(-n))
h_N_27 <- Linv %*% h_N + f(0)*ynew
theta_N_27 <- solve(F_N_27)%*%h_N_27
```

$$\begin{aligned} \boldsymbol{F}_{N+1} &= \boldsymbol{F}_N + \boldsymbol{f}(-N)\boldsymbol{f}^T(-N) \\ \boldsymbol{h}_{N+1} &= \boldsymbol{L}^{-1}\boldsymbol{h}_N + \boldsymbol{f}(0)\,Y_{N+1} \\ \widehat{\boldsymbol{\theta}}_{N+1} &= \boldsymbol{F}_{N+1}^{-1}\boldsymbol{h}_{N+1} \end{aligned}$$

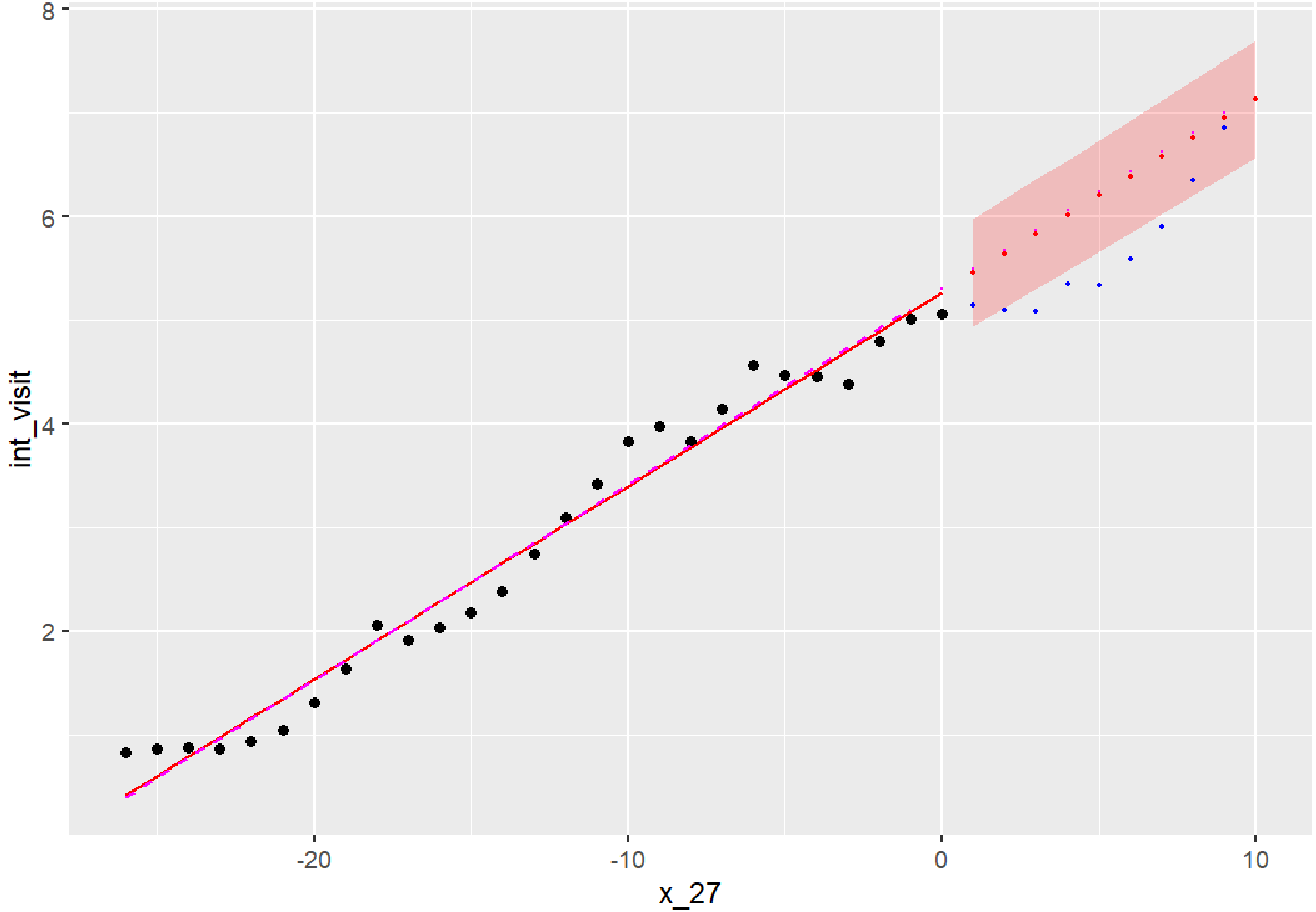
```
> print(theta_N_27)
       [,1]
  5.2702522
j 0.1867399
```

Slight change from before:

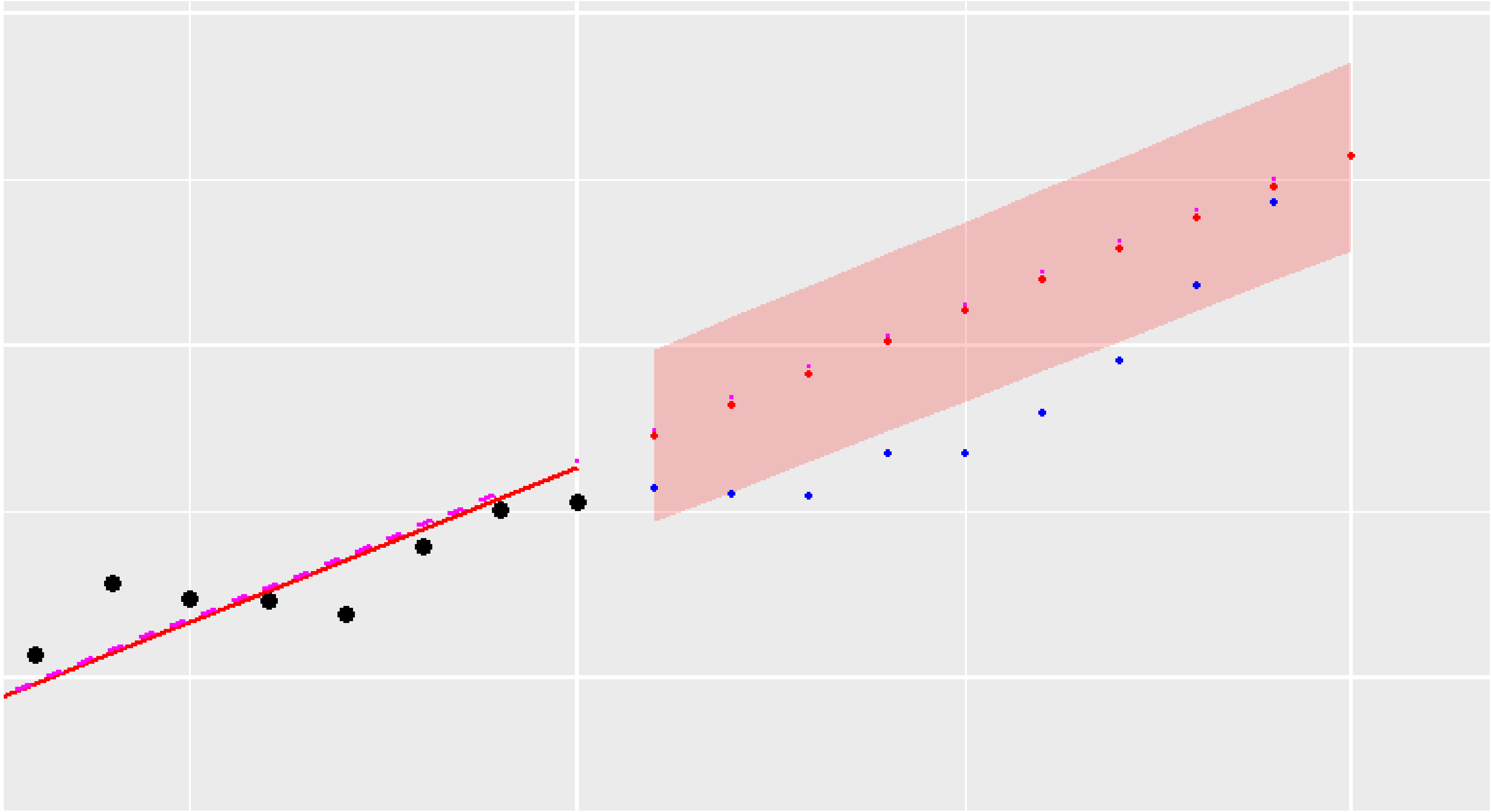
5.11

0.19

# Example in R



# Example in R



# Outline of the lecture

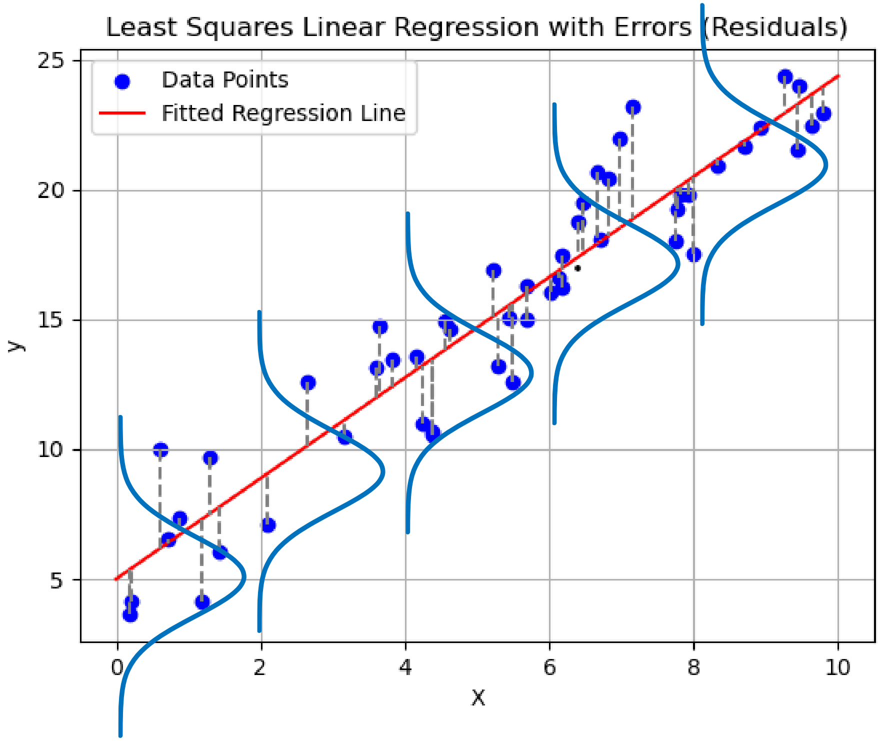
- Regression based methods, 1st part:
  - Ordinary Least Squares (OLS)
  - Predictions = forecast
  - Global Trend Model
  - **Weighted Least Squares (WLS)**

# Weighted Least Squares (WLS) estimates

Recall: OLS, minimizing the sum of squared residuals:

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta})$$

Where we assumed the residuals have the same variance and are mutually uncorrelated.

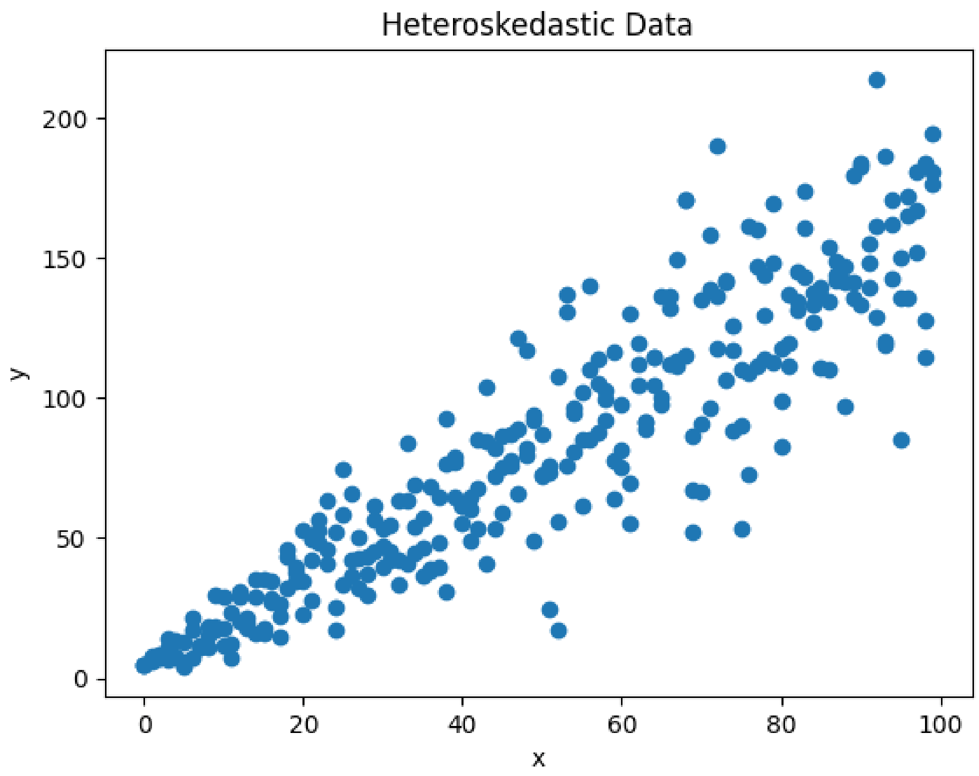


# Weighted Least Squares (WLS) estimates

Recall: OLS, minimizing the sum of squared residuals:

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta})$$

Where we assumed the residuals have the same variance and are mutually uncorrelated.



# Weighted Least Squares (WLS) estimates

Recall: OLS, minimizing the sum of squared residuals:

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta})$$

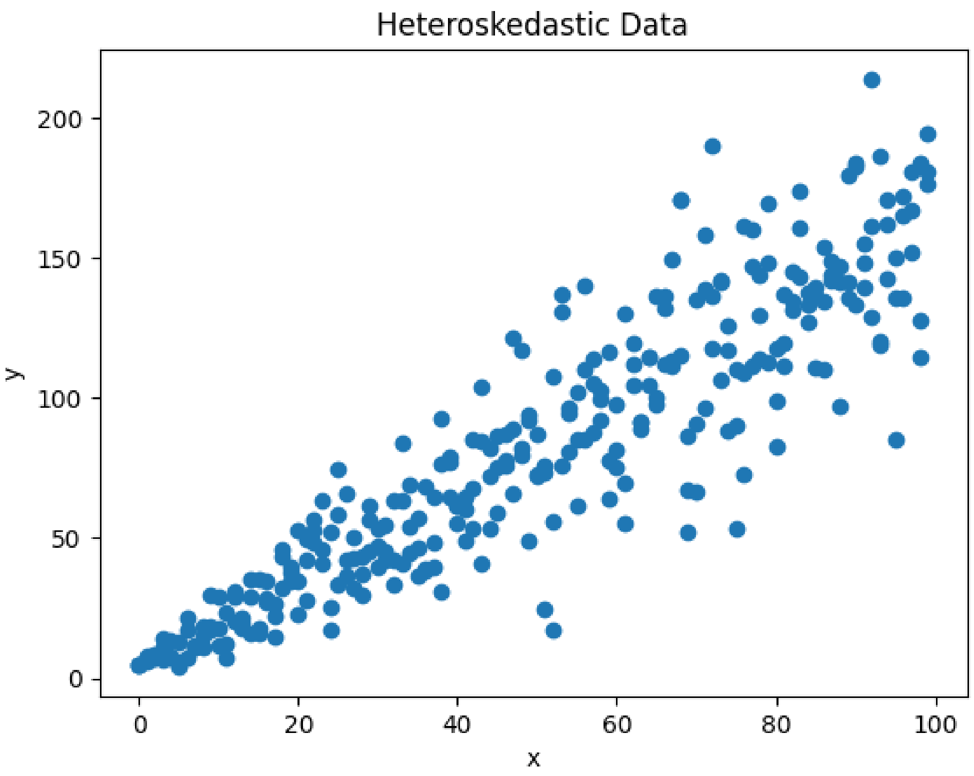
Where we assumed the residuals have the same variance and are mutually uncorrelated.

In WLS we assume the residuals can have different variances and be mutually correlated:

$$V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \boldsymbol{\Sigma}$$

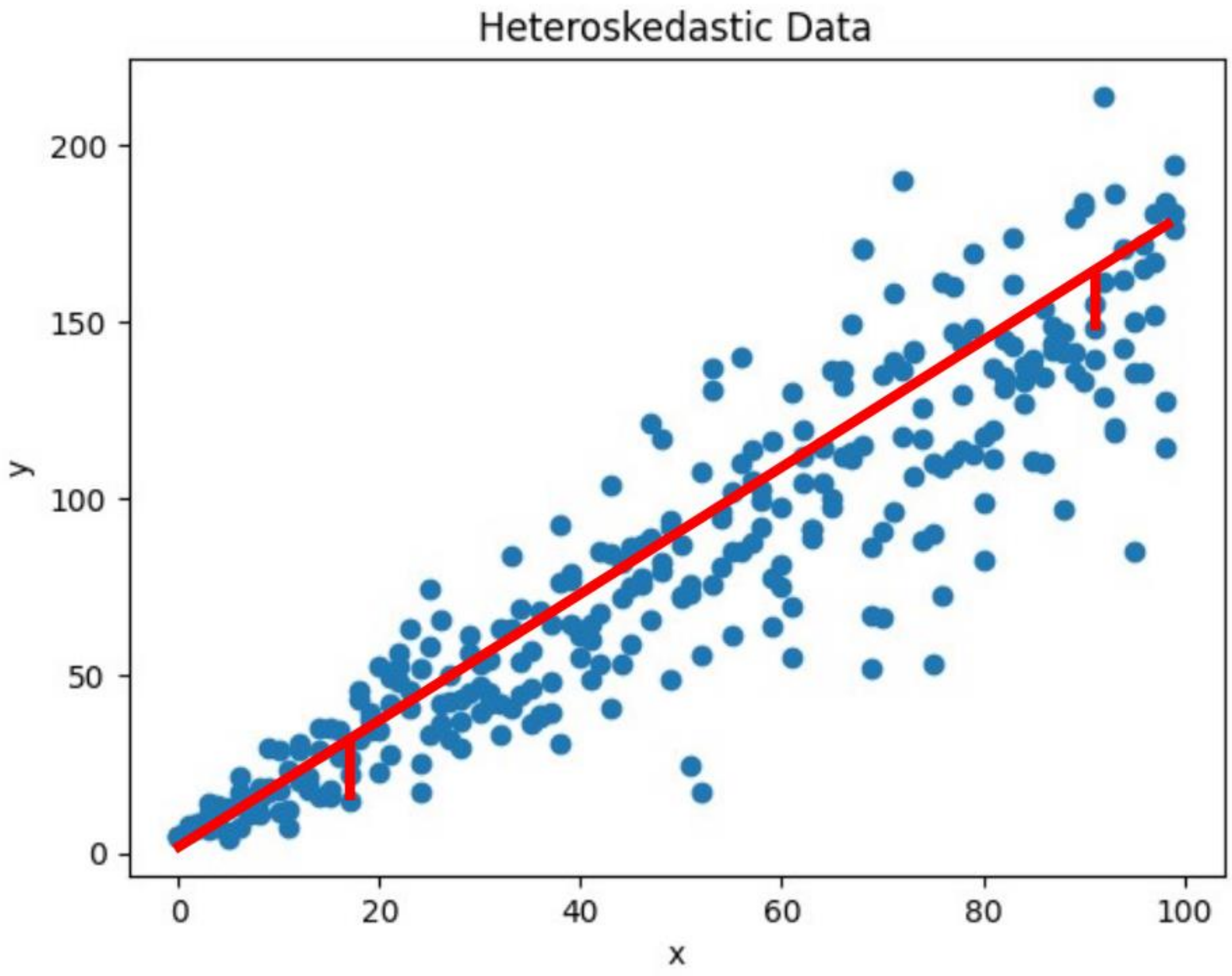
We minimize the *weighted* sum of squared residuals:

$$(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})$$



# Weighted Least Squares (WLS) estimates

How would weighting residuals make sense for this data?



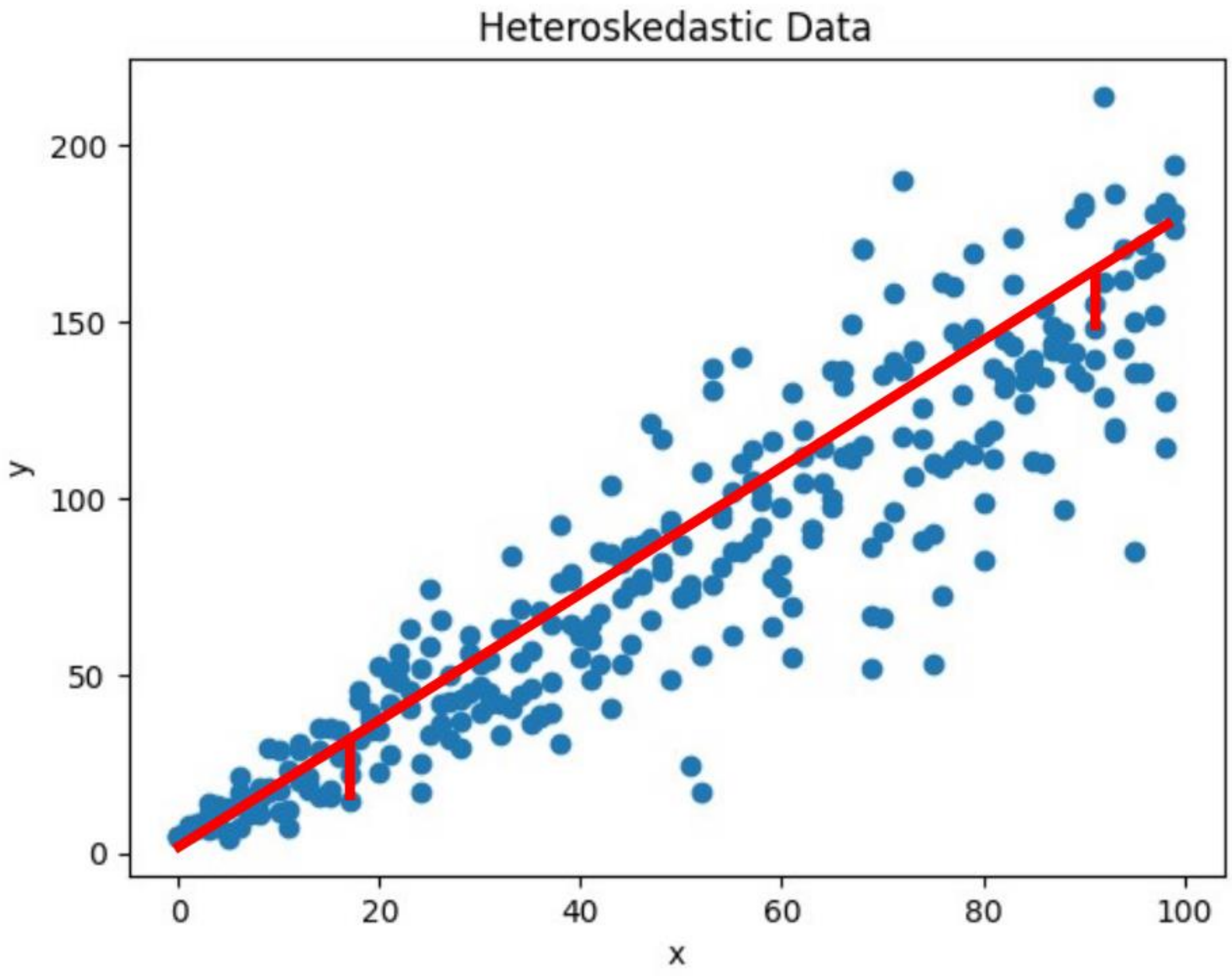
# Weighted Least Squares (WLS) estimates

How would weighting residuals make sense for this data?

Large variance

↓

Lower ”weight”



# WLS estimates for the general linear model

For all observations the model equations are written as:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad or \quad \boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

We want to minimize $(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\theta})$

The solution is

$$\boxed{\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x})^{-1} \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}}$$

(if $\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$ is invertible)

$$\widehat{\sigma}^2 = \frac{1}{n-p} (\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}}) \qquad V[\widehat{\boldsymbol{\theta}}] = E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] = (\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x})^{-1} \sigma^2$$

# Maximum likelihood estimates

- Assume that the observations are Gaussian;

$$\boldsymbol{Y} \sim \mathrm{N}_n(\boldsymbol{x\theta}, \sigma^2\boldsymbol{\Sigma})$$

- and that $\boldsymbol{\Sigma}$ is known.
- The ML-estimator is (here) the same as the WLS-estimator:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

- The ML-estimator for $\sigma^2$ is

$$\widehat{\sigma}^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\widehat{\boldsymbol{\theta}})$$

- OLS and WLS estimates can be interpreted as? assumptions of Gaussianity.

# WLS variance-covariance examples

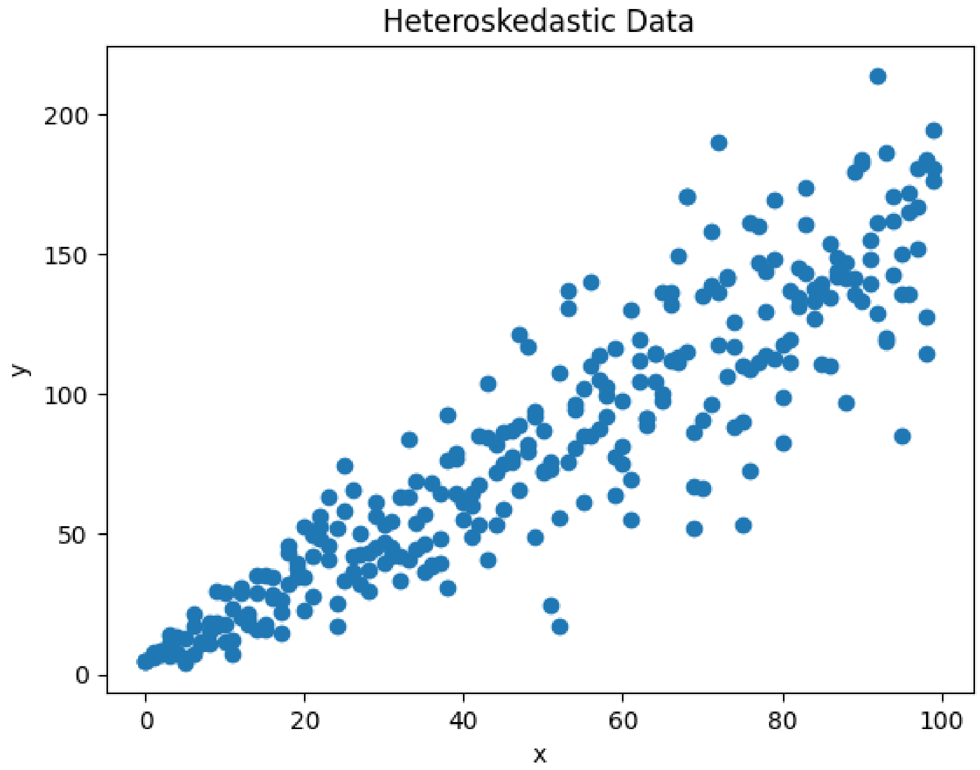
$$V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2\boldsymbol{\Sigma}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \cdots & \sigma_{nn} \end{bmatrix}$$

# WLS variance-covariance examples

$$V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \boldsymbol{\Sigma}$$

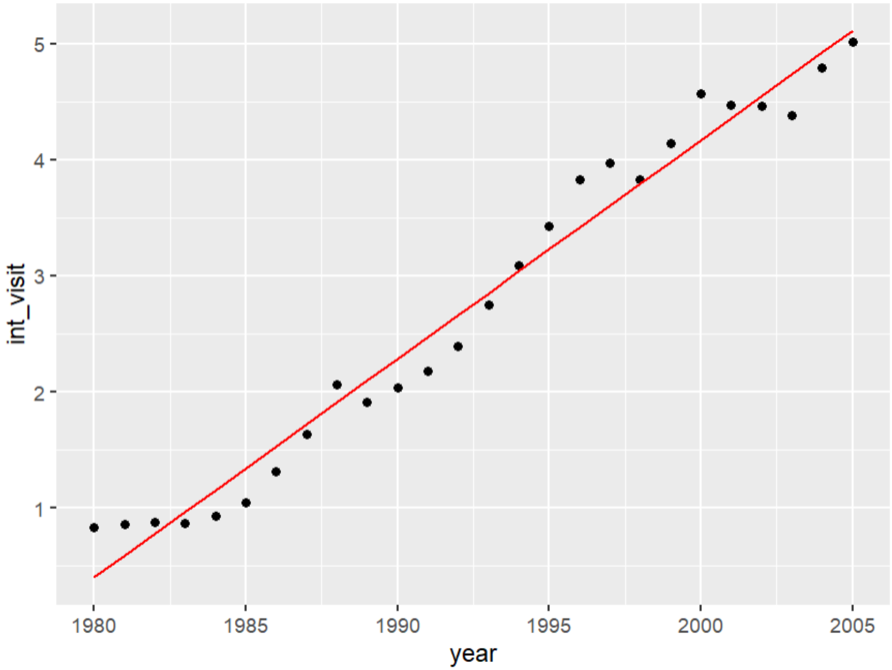
$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{22} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{nn} \end{bmatrix}$$



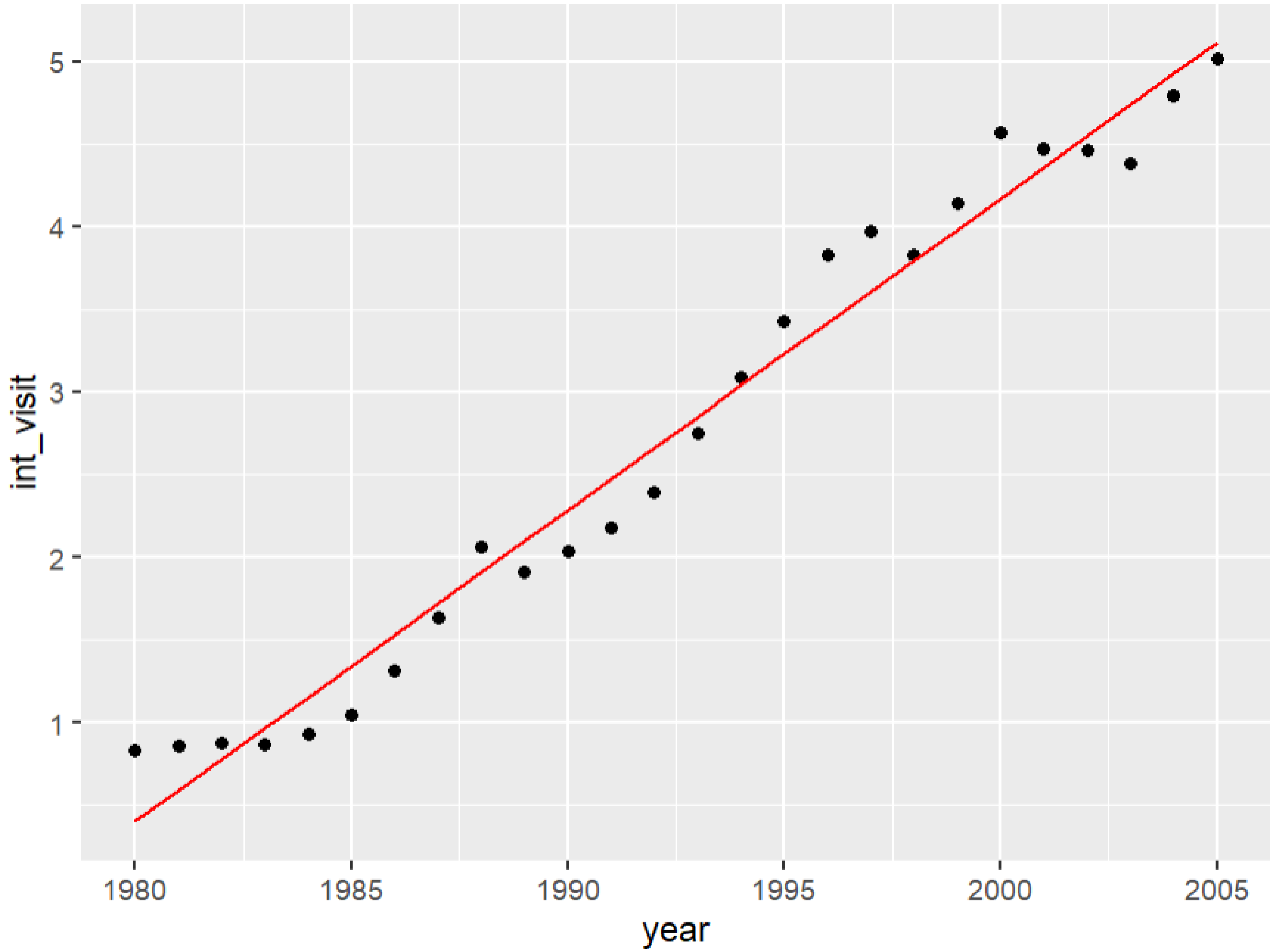
# WLS variance-covariance examples

$$V[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2\boldsymbol{\Sigma}$$

$$\Sigma = \begin{bmatrix} 1 & \rho^1 & \rho^2 & \dots & \rho^n \\ \rho^1 & 1 & \rho^1 & \dots & \rho^{n-1} \\ \rho^2 & \rho^1 & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^n & \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$



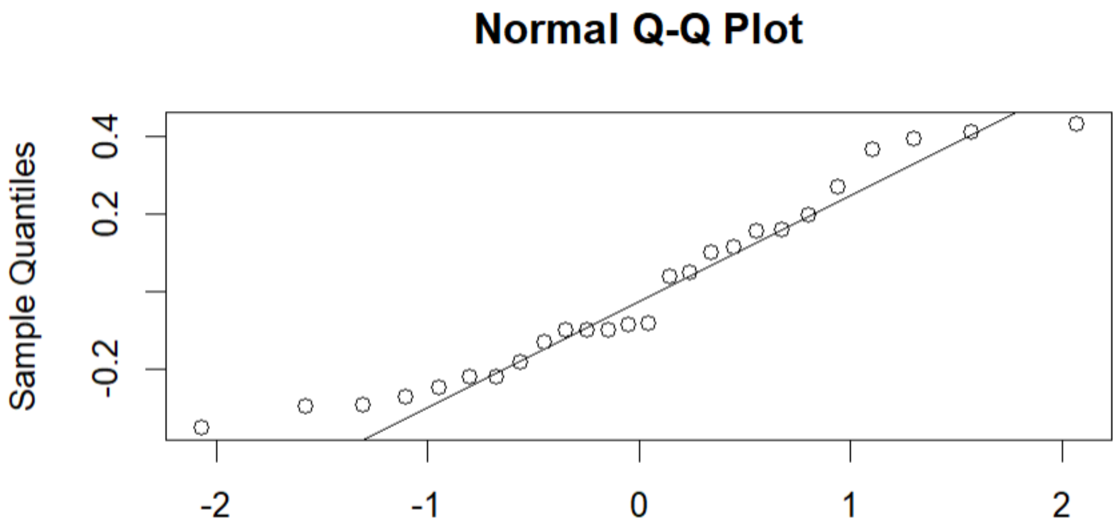
# Example in R



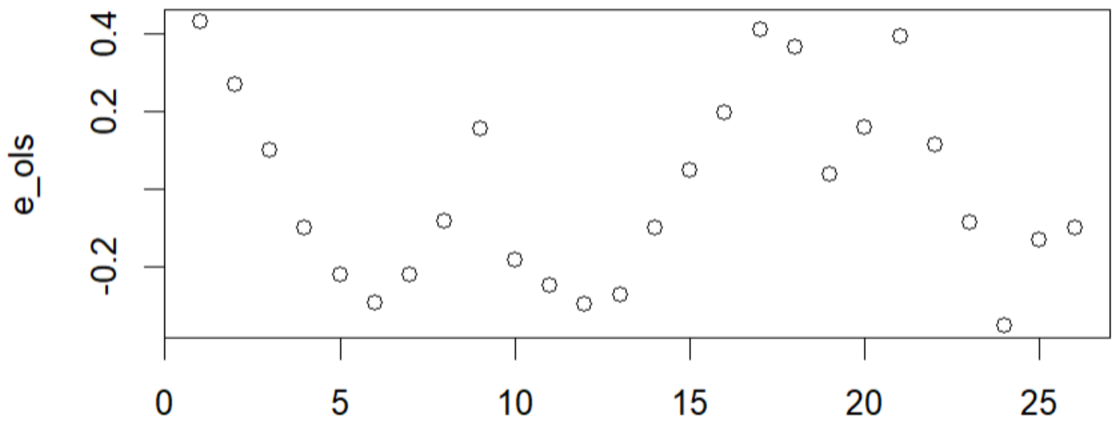

Is it a good model?

We should inspect the residuals!

# Example in R



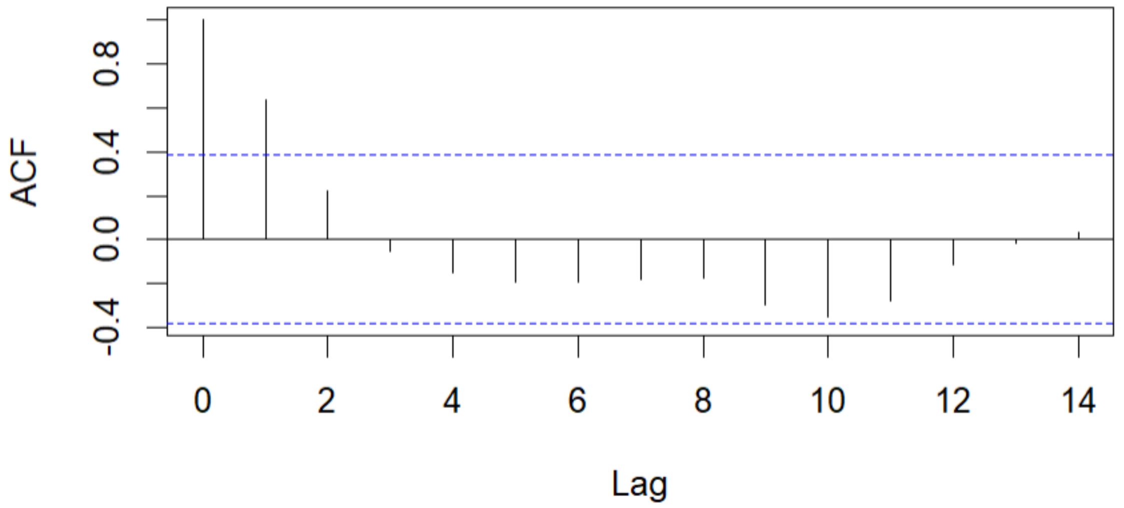

Are the residuals normally distributed?

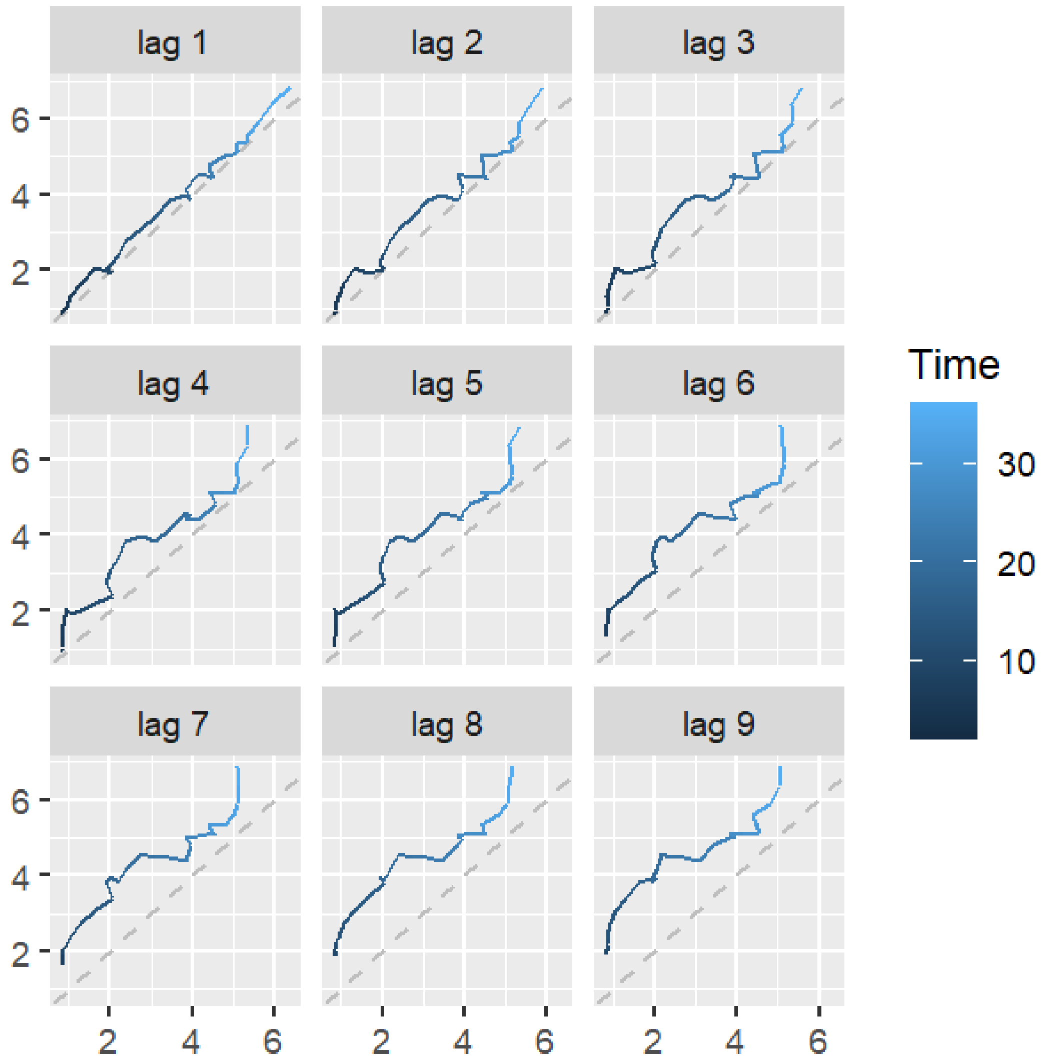



Is there a pattern in the residuals?

# Example in R

We can also inspect the autocorrelations of the resilduals



# Example in R

```
> print(SIGMA[1:5,1:5])
       [,1]  [,2] [,3]  [,4]   [,5]
[1,] 1.0000 0.700 0.49 0.343 0.2401
[2,] 0.7000 1.000 0.70 0.490 0.3430
[3,] 0.4900 0.700 1.00 0.700 0.4900
[4,] 0.3430 0.490 0.70 1.000 0.7000
[5,] 0.2401 0.343 0.49 0.700 1.0000
```

```
WLS <- solve(t(X)%*%solve(SIGMA)%*%X)%*%(t(X)%*%solve(SIGMA)%*%y)
```

```
print(WLS)
# [1,] -356.7172120
# [2,]    0.1804264
```

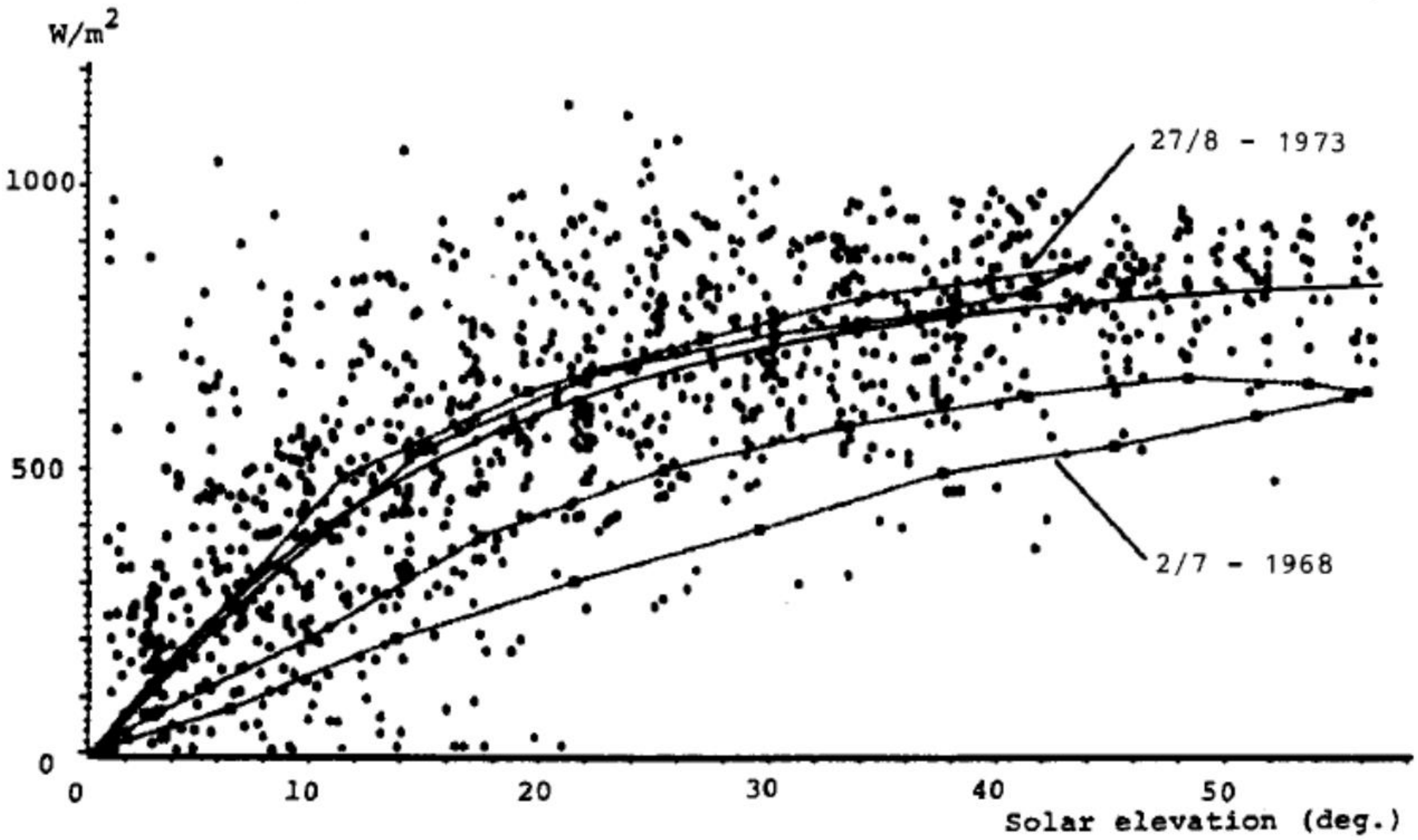
The estimated parameters are slightly different

# Example in R


Blue is WLS

Red is OLS from before

# Example from the book




Variance is larger for small angles

AND

Observations are correlated in time

$$\Sigma_{ij} = \frac{\rho^{|t_i - t_j|}}{\sin(h_{t_i})\sin(h_{t_j})}$$

# Next time

- Combine weights (WLS) with trend models to make ”local trend models”

- ”Exponential smoothing”