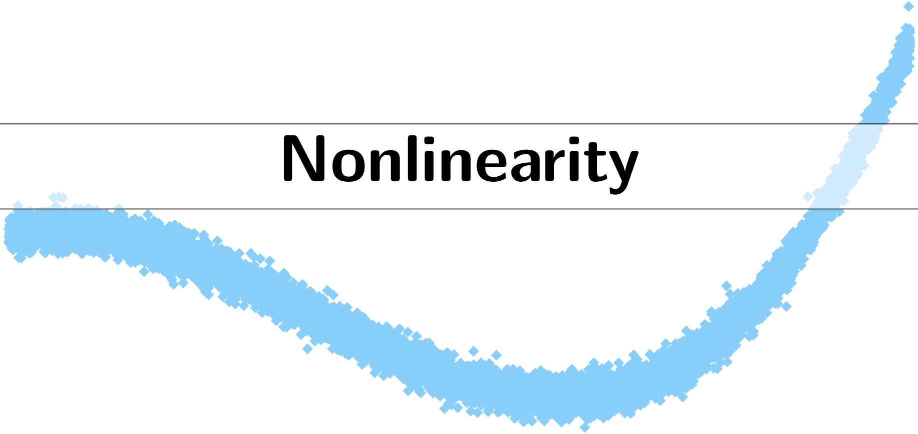
Nonlinearity

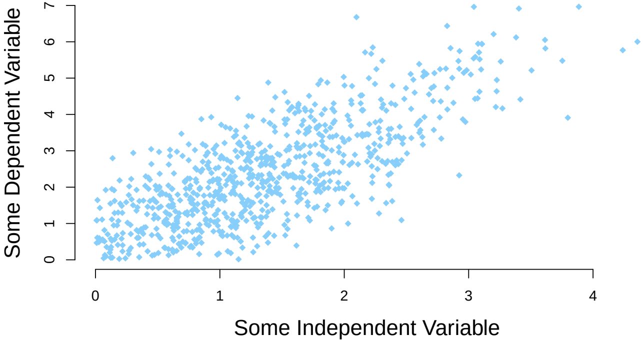


Today:

- Understand some examples of what nonlinearity can look like.
- Apply two methods for dealing with nonlinearity.

What does nonlinearity look like?

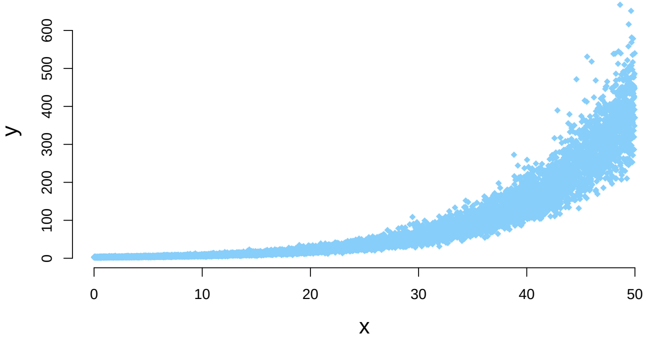
- Linear regression allows us to model a linear relationship between x and y;



What does nonlinearity look like?

- Linear regression allows us to model a linear relationship between x and y;

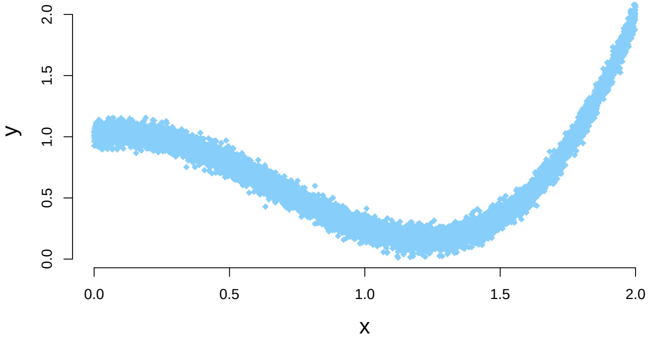
- But what if the data looks like this?!



What does nonlinearity look like?

- Linear regression allows us to model a linear relationship between x and y;

- But what if the data looks like this?!



Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. To see this first think about 'normal' linear regression:

$$y_1 = \beta_0 + \beta_1 x$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. To see this first think about 'normal' linear regression:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x \\ \text{for a unit increase in } x\text{: } y_2 &= \beta_0 + \beta_1(x + 1) \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. To see this first think about 'normal' linear regression:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x \\ \text{for a unit increase in } x\text{: } y_2 &= \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1 \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. To see this first think about 'normal' linear regression:

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x \\
\text{for a unit increase in } x\text{: } y_2 &= \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1 \\
\text{so change in } y \text{ is: } y_2 - y_1 &
\end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. To see this first think about 'normal' linear regression:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x \\ \text{for a unit increase in } x\text{: } y_2 &= \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1 \\ \text{so change in } y \text{ is: } y_2 - y_1 &= \underbrace{\beta_0 + \beta_1 x + \beta_1}_{y_2} - \underbrace{\beta_0 - \beta_1 x}_{y_1} = \beta_1. \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y…

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\ln\{y_1\} = \beta_0 + \beta_1 x$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\begin{aligned} \ln\{y_1\} &= \beta_0 + \beta_1 x \\ \text{for a unit increase in } x\text{:}\quad \ln\{y_2\} &= \beta_0 + \beta_1 (x+1) \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\begin{aligned} \ln\{y_1\} &= \beta_0 + \beta_1 x \\ \text{for a unit increase in } x\text{: } \ln\{y_2\} &= \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1 \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\exp\{\ln\{y_1\}\} = \exp\{\beta_0 + \beta_1 x\}$$

$$\text{for a unit increase in } x\text{: } \exp\{\ln\{y_2\}\} = \exp\{\beta_0 + \beta_1(x+1)\} = \exp\{\beta_0 + \beta_1 x + \beta_1\}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\begin{aligned} y_1 &= \exp\{\beta_0 + \beta_1 x\} \\ \text{for a unit increase in } x\text{: } y_2 &= \exp\{\beta_0 + \beta_1 (x+1)\} = \exp\{\beta_0 + \beta_1 x + \beta_1\} \end{aligned}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$y_1 = \exp\{\beta_0 + \beta_1 x\}$$

$$\text{for a unit increase in } x\text{: } y_2 = \exp\{\beta_0 + \beta_1(x+1)\} = \exp\{\beta_1\}\exp\{\beta_0 + \beta_1 x\}$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$y_1 = \exp\{\beta_0 + \beta_1 x\}$$

$$\text{for a unit increase in } x\text{: } y_2 = \exp\{\beta_0 + \beta_1(x)\} = \exp\{\beta_1\}\exp\{\beta_0 + \beta_1 x)\}$$

$$\text{so change in } y \text{ is: } y_2 - y_1$$

Strategy 1 for dealing with nonlinearity

- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\begin{aligned} y_1 &= \exp\{\beta_0 + \beta_1 x\} \\ \text{for a unit increase in } x\text{: } y_2 &= \exp\{\beta_0 + \beta_1 (x)\} = \exp\{\beta_1\}\exp\{\beta_0 + \beta_1 x)\} \\ \text{so change in } y \text{ is: } y_2 - y_1 &= \underbrace{\exp\{\beta_1\}\exp\{\beta_0 + \beta_1 x)\}}_{y_2} - \underbrace{\exp\{\beta_0 + \beta_1 x\}}_{y_1} \end{aligned}$$

Strategy 1 for dealing with nonlinearity

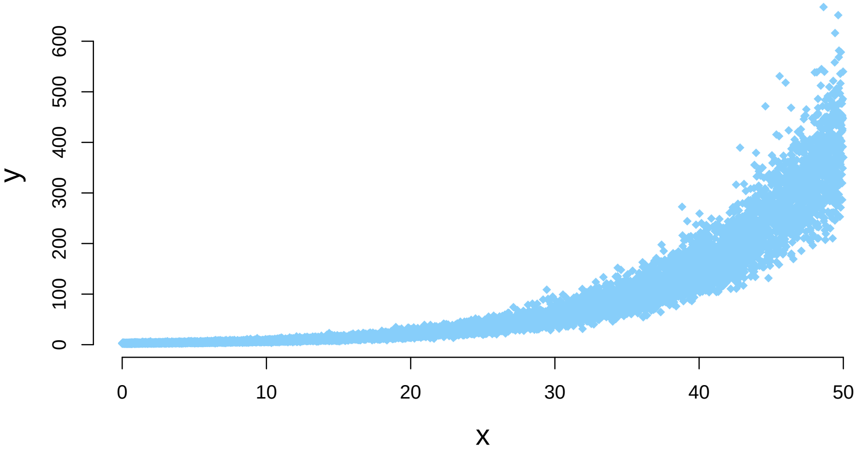
- Modify the dependent variable y to turn the nonlinear problem into a linear problem;
- Very typical to natural log y...

$$\text{i.e. instead of } y = \beta_0 + \beta_1 x + \varepsilon \text{ we use: } \ln\{y\} = \beta_0 + \beta_1 x + \varepsilon;$$

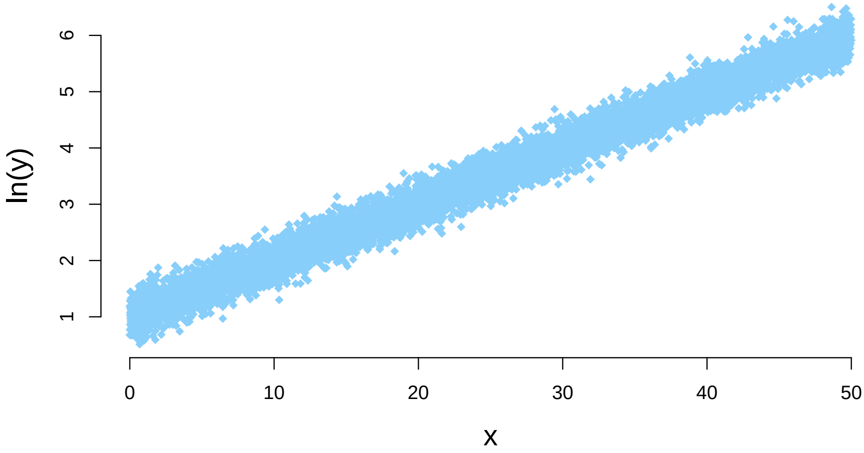
- Apply linear regression to estimate β_0 and β_1;
- Note – if you do this you need to be careful with the interpretation of the βs. Now think about 'logged' linear regression:

$$\begin{aligned} y_1 &= \exp\{\beta_0 + \beta_1 x\} \\ \text{for a unit increase in } x\text{: } y_2 &= \exp\{\beta_0 + \beta_1(x)\} = \exp\{\beta_1\}\exp\{\beta_0 + \beta_1 x)\} \\ \text{so change in } y \text{ is: } y_2 - y_1 &= (\exp\{\beta_1\} - 1)\exp\{\beta_0 + \beta_1 x)\}. \end{aligned}$$

Strategy 1 for dealing with nonlinearity



Strategy 1 for dealing with nonlinearity



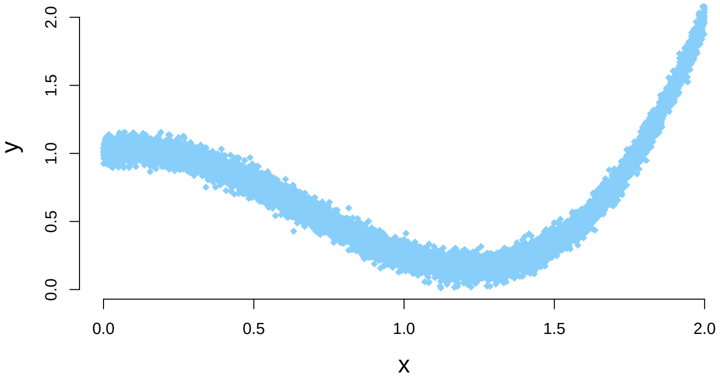
Strategy 2 for dealing with nonlinearity

- Modify the independent variable x to model the nonlinearity;
- Very typical to use a polynomial of x, for example x^2 or x^3:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

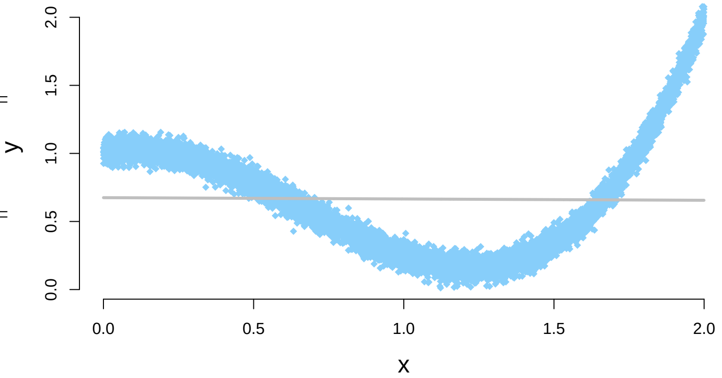
- Apply linear regression to estimate βs;
- Note – if you do this do not extrapolate beyond the boundary of the sample.

Strategy 2 for dealing with nonlinearity



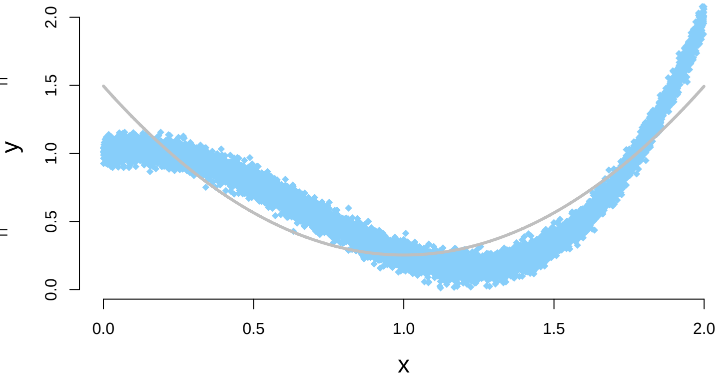
What about a simple **linear** model?

	Coefficient	*p*-value
Incpt	0.6750	0.0
x	-0.0096	0.185



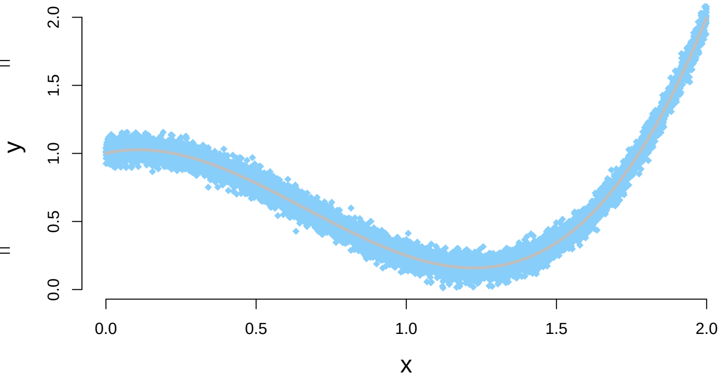
What about a **quadratic** model?

	Coefficient	p-value
Incpt	1.4952	0.0
x	-2.4825	0.0
x^2	1.2409	0.0



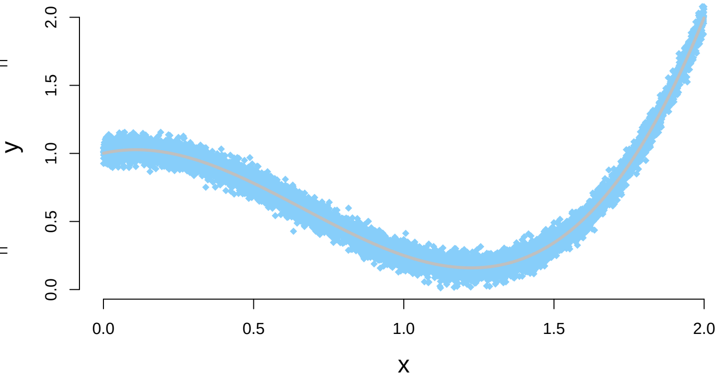
What about a **cubic polynomial** model?

	Coefficient	p-value
Incpt	1.0012	0.0
x	0.4934	0.0
x^2	-2.4914	0.0
x^3	1.2469	0.0



Here is the true model: $y = 1.0 + 0.5x - 2.5x^2 + 1.25x^3 + \varepsilon$.

	Coefficient	p-value
Incpt	1.0	0.0
x	0.5	0.0
x^2	-2.5	0.0
x^3	1.25	0.0



Examples...

- What is the optimal top marginal tax rate for economic growth?
- Does inequality make democracy more or less likely?

Why should we care?

Linear regression can deal with non-linear situations – this broadens its applicability and makes it even more powerful.