All Things Regression

Part II

Fernando Hoces la Guardia

07/21/2022

Housekeeping

- Let's choose the chapter for the summary (still due Friday 5pm on gradescope)
- Practice questions are up. Midterm will follow similar questions (but not exactly the same ones).
  - Goal: if you understood the concepts behind the practice questions, you will do well in the midterm.
- Switching the order of the review session: will do a review on Monday (before the midterm), and on Wednesday we will start with new material. Bring questions! (I will not bring new material, if we finish early we can watch the first part of Run Lola Run)

Regression Journey

- Regression as Matching on Groups. Ch2 of MM up to page 68 (not included).
- Regression as Line Fitting and Conditional Expectation. Ch2 of MM, Appendix.
- Multiple Regression and Omitted Variable Bias. Ch2 of MM pages 68-79 and Appendix.
- All Things Regression: Anatomy, Inference, Logarithms, Binary Outcomes, and R^2. Ch2 of MM, Appendix + others.

Regression Journey

- Regression as Matching on Groups. Ch2 of MM up to page 68 (not included).
- Regression as Line Fitting and Conditional Expectation. Ch2 of MM, Appendix.
- Multiple Regression and Omitted Variable Bias. Ch2 of MM pages 68-79 and Appendix.
- **All Things Regression: Anatomy, Inference, Logarithms, Binary Outcomes, and R^2. Ch2 of MM, Appendix + others.**

Today and Tomorrow's Lecture

- Regression Anatomy

- Regression Inference

- R^2

- Non-linearities:

    - Logarithms
    - Others

- Binary Outcomes

Analysis of Variance

- Remember that $Y_i = \widehat{Y_i} + e_i$.

- We have the following decomposition

$$
\begin{aligned}
Var(Y) &= Var(\widehat{Y} + e) \\
&= Var(\widehat{Y}) + Var(e) + 2Cov(\widehat{Y}, e) \\
&= Var(\widehat{Y}) + Var(e)
\end{aligned}
$$

- **Total variation (SST) = Model explained (SSE) + Unexplained (SSR)**

- Because:

  - $Var(x + y) = Var(x) + Var(y) + 2Cov(x, y)$
  - $Cov(\hat{Y}, e) = 0$

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

$$R^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1]$$

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

$$R^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1]$$

- R^2 close to 1 indicates a **very *high* explanatory power** of the model.

- R^2 close to 0 indicates a **very *low* explanatory power** of the model.

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

$$R^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1]$$

- R^2 close to 1 indicates a **very *high* explanatory power** of the model.
- R^2 close to 0 indicates a **very *low* explanatory power** of the model.
- *Interpretation:* an R^2 of 0.5, for example, means that the variation in x "explains" 50% of the variation in y.

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

$$R^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1]$$

- R^2 close to 1 indicates a **very *high* explanatory power** of the model.
- R^2 close to 0 indicates a **very *low* explanatory power** of the model.
- *Interpretation:* an R^2 of 0.5, for example, means that the variation in x "explains" 50% of the variation in y.
- ⚠️ Low R^2 does **NOT** mean it's a useless model! Remember that econometrics is interested in causal mechanisms, not prediction!

Goodness of Fit

- The R^2 measures how well the **model fits the data**.

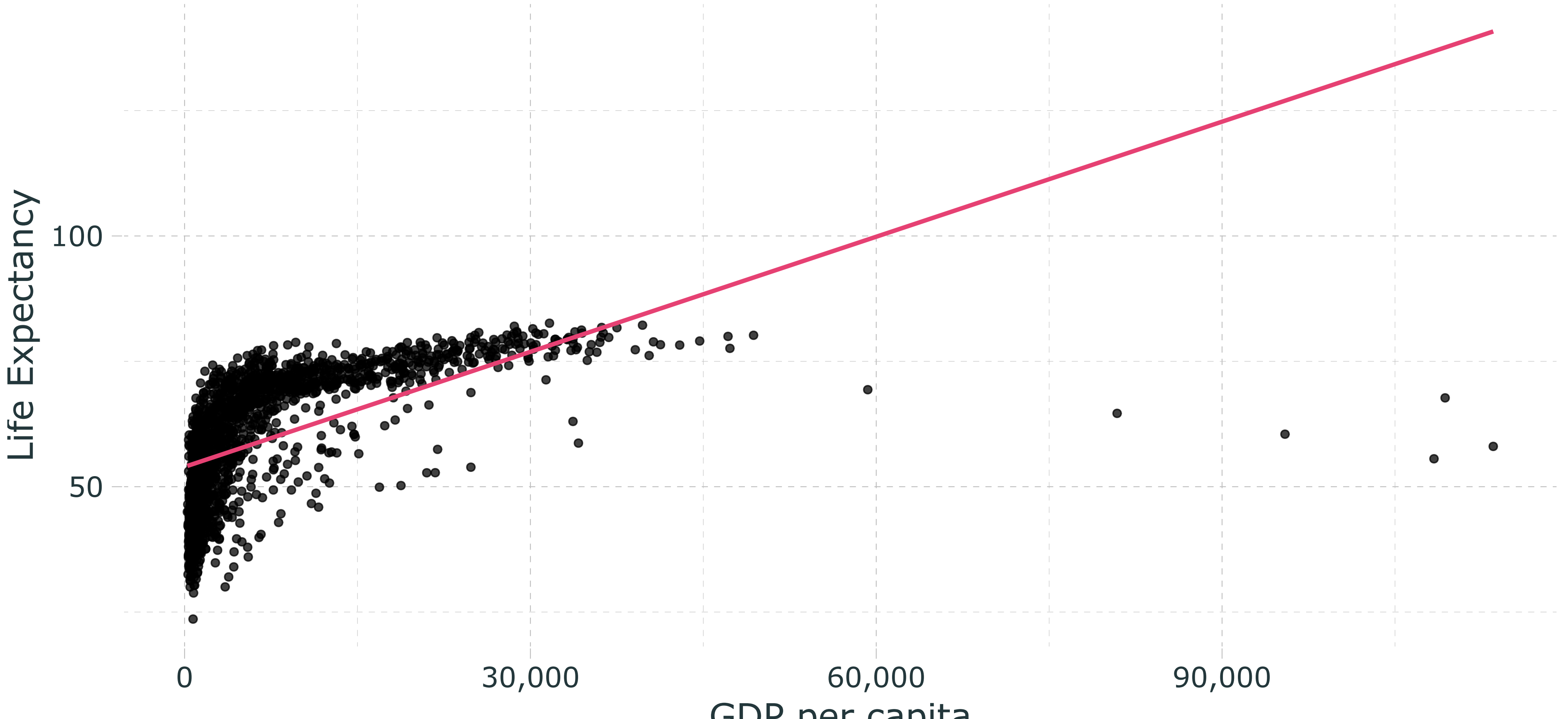
$$R^2 = \frac{\text{variance explained}}{\text{total variance}} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1]$$

- R^2 close to 1 indicates a **very *high* explanatory power** of the model.
- R^2 close to 0 indicates a **very *low* explanatory power** of the model.
- *Interpretation:* an R^2 of 0.5, for example, means that the variation in x "explains" 50% of the variation in y.
- ⚠️ Low R^2 does **NOT** mean it's a useless model! Remember that econometrics is interested in causal mechanisms, not prediction!
- ⚠️ The R^2 is **NOT** an indicator of whether a relationship is causal!

Non-linearities

Non-linearities

$$\widehat{(\text{Life Expectancy})}_i = 53.96 + 8 \times 10^{-4} \cdot \text{GDP}_i$$



Nonlinear Relationships

Erroneus critique of regression: "many economic relationships are **nonlinear** (*e.g.*, most production functions, profit, diminishing marginal utility, tax revenue as a function of the tax rate, *etc.*), hence fitting straight lines is a bad way of estimating such relationships"

Nonlinear Relationships

Erroneus critique of regression: "many economic relationships are **nonlinear** (*e.g.*, most production functions, profit, diminishing marginal utility, tax revenue as a function of the tax rate, *etc.*), hence fitting straight lines is a bad way of estimating such relationships"

The flexibility of regression OLS estimation can accommodate many, but not all, nonlinear relationships.

- Underlying model must be linear-in-parameters.
- Nonlinear transformations of variables are okay.

Linearity

Linear-in-parameters: Parameters enter model as a weighted sum, where the weights are functions of the variables.

- This is the one required to estimate OLS

Linear-in-variables: Variables enter the model as a weighted sum, where the weights are functions of the parameters.

- This is the one the critique was pointing at. OLS works perfectly here.

Linearity

Linear-in-parameters: Parameters enter model as a weighted sum, where the weights are functions of the variables.

- This is the one required to estimate OLS

Linear-in-variables: Variables enter the model as a weighted sum, where the weights are functions of the parameters.

- This is the one the critique was pointing at. OLS works perfectly here.

The standard linear regression model satisfies both properties:

$$\boldsymbol{Y_i} = \beta_0 + \beta_1 \boldsymbol{X_{1i}} + \beta_2 \boldsymbol{X_{2i}} + \cdots + \beta_k \boldsymbol{X_{ki}} + \boldsymbol{e_i}$$

Linearity

Which of the following is **linear-in-parameters**, **linear-in-variables**, or **neither**?

1. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + e_i$
2. $Y_i = \beta_0 X_i^{\beta_1} e_i$
3. $Y_i = \beta_0 + \beta_1 \beta_2 X_i + e_i$

Linearity

Which of the following is **linear-in-parameters**, **linear-in-variables**, or **neither**?

1. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + e_i$

2. $Y_i = \beta_0 X_i^{\beta_1} e_i$

3. $Y_i = \beta_0 + \beta_1 \beta_2 X_i + e_i$

Model 1 is linear-in-parameters, but not linear-in-variables.

Linearity

Which of the following is **linear-in-parameters**, **linear-in-variables**, or **neither**?

1. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + e_i$

2. $Y_i = \beta_0 X_i^{\beta_1} e_i$

3. $Y_i = \beta_0 + \beta_1 \beta_2 X_i + e_i$

Model 1 is linear-in-parameters, but not linear-in-variables.

Model 2 is neither.

Linearity

Which of the following is **linear-in-parameters**, **linear-in-variables**, or **neither**?

1. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + e_i$

2. $Y_i = \beta_0 X_i^{\beta_1} e_i$

3. $Y_i = \beta_0 + \beta_1 \beta_2 X_i + e_i$

Model 1 is linear-in-parameters, but not linear-in-variables.

Model 2 is neither.

Model 3 is linear-in-variables, but not linear-in-parameters.

We're Going to Take Logs

The natural log is the inverse function for the exponential function:

$\log(e^x) = x$ for $x > 0$.

(Natural) Log Rules and Approximations

1. Product rule: $\log(AB) = \log(A) + \log(B)$.

2. Quotient rule: $\log(A/B) = \log(A) - \log(B)$.

3. Power rule: $\log(A^B) = B \cdot \log(A)$.

4. $\log(e) = 1$, $\log(1) = 0$, and $\log(x)$ is undefined for $x \leq 0$.

5. **Approximation:** $\log(1 + A) = A$ **If A is very small (~less than 0.2)**

Log-Linear Model

Nonlinear Model

$$Y_i = \alpha e^{\beta_1 X_i} e_i$$

- $Y > 0$, X is continuous, and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$\log(Y_i) = \log(\alpha) + \beta_1 X_i + \log(e_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(e_i) \equiv e_i$.

Transformed (Linear) Model

$$\log(Y_i) = \beta_0 + \beta_1 X_i + e_i$$

- *Can* estimate with OLS, but coefficient interpretation changes.

Log-Linear Model

$$\log(Y_i) = \beta_0 + \beta_1 X_i + e_i$$

Interpretation

- A one-unit increase in the explanatory variable increases the outcome variable by approximately $\beta_1 \times 100$ percent, on average.
- *Example:* If $\log(\hat{\text{Pay}}_i) = 2.9 + 0.03 \cdot \text{School}_i$, then an additional year of schooling increases pay by approximately 3 percent, on average.

Why?

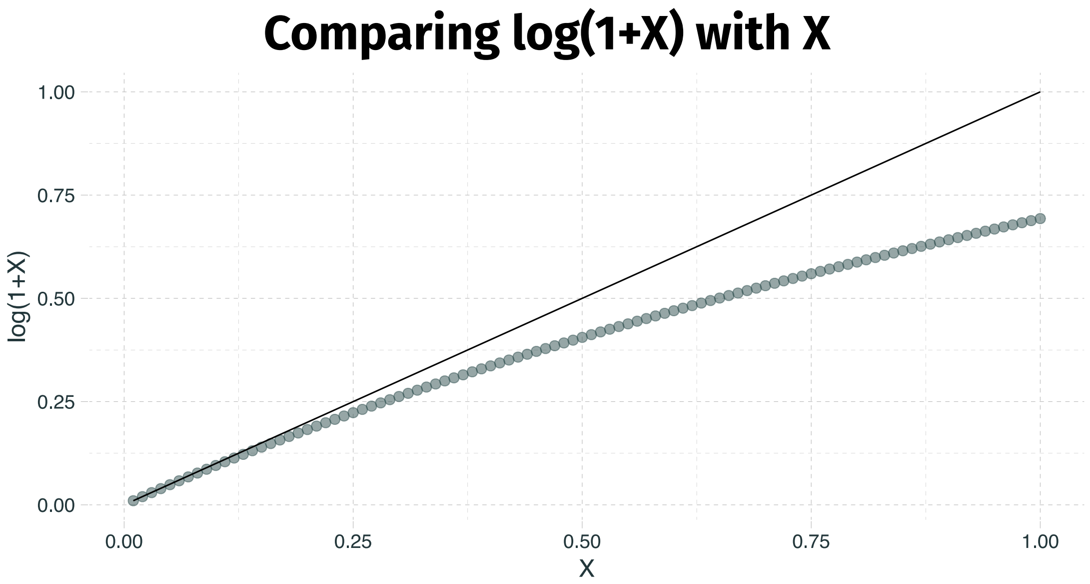
- We want to know how to interpret what is the associated increase in Y, when we increase X in one unit.

$$\log(Y_i) = \beta_0 + \beta_1 X_i + e_i$$
$$\widetilde{\log}(Y_i) = \beta_0 + \beta_1 (X_i + 1) + e_i$$

Why?

- We want to know how to interpret what is the associated increase in Y, when we increase X in one unit.

$$
\begin{aligned}
\log(Y_i) &= \beta_0 + \beta_1 X_i + e_i \\
\widetilde{\log}(Y_i) &= \beta_0 + \beta_1 (X_i + 1) + e_i \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_0 + \beta_1 X_i + \beta_1 + e_i - \\
&\qquad (\beta_0 + \beta_1 X_i + e_i) \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_1
\end{aligned}
$$

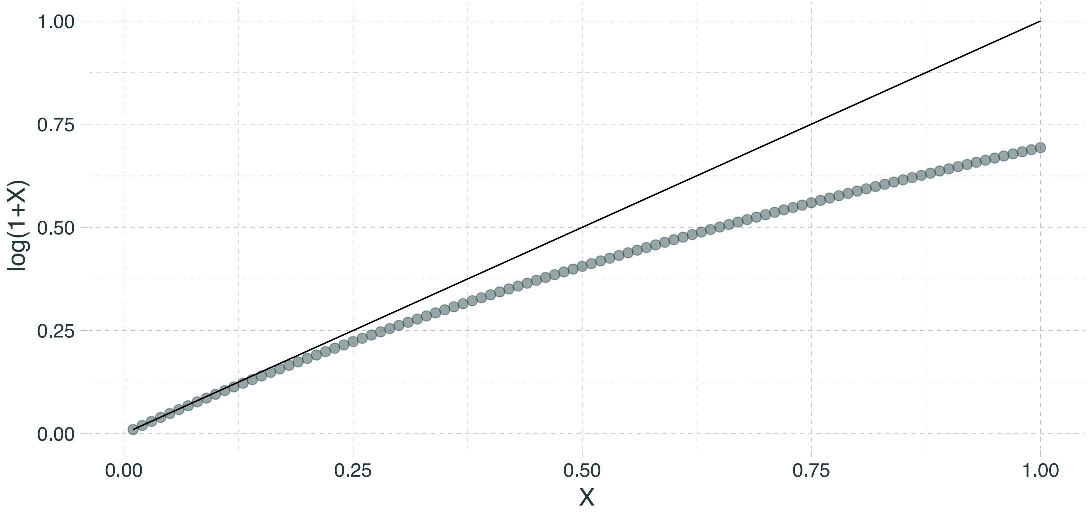


Why?

- We want to know how to interpret what is the associated increase in Y, when we increase X in one unit.

$$
\begin{aligned}
\log(Y_i) &= \beta_0 + \beta_1 X_i + e_i \\
\widetilde{\log}(Y_i) &= \beta_0 + \beta_1 (X_i + 1) + e_i \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_0 + \beta_1 X_i + \beta_1 + e_i - \\
&\qquad (\beta_0 + \beta_1 X_i + e_i) \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_1 \\
\widetilde{\log}(Y_i) &= \log(Y_i) + \beta_1 \\
\widetilde{\log}(Y_i) &\approx \log(Y_i) + log(1 + \beta_1) \\
\widetilde{\log}(Y_i) &\approx \log(Y_i(1 + \beta_1)) \\
\widetilde{Y}_i &\approx Y_i(1 + \beta_1)
\end{aligned}
$$

Comparing log(1+X) with X



- A one-unit increase in the explanatory variable increases the outcome variable by approximately $\beta_1 \times 100$ percent, on average.
- What if β_1 is large (>0.2)? No problem, just divide X by 10, 100, or larger, to shrink the units of β_1.

(If X is Binary and $\beta > 0.2$: Use Exact)

$$
\begin{aligned}
\log(Y_i) &= \beta_0 + \beta_1 X_i + e_i \\
\widetilde{\log}(Y_i) &= \beta_0 + \beta_1 (X_i + 1) + e_i \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_0 + \beta_1 X_i + \beta_1 + e_i - \\
&\quad (\beta_0 + \beta_1 X_i + e_i)
\end{aligned}
$$

Exact

$$
\begin{aligned}
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_1 \\
\log(\widetilde{Y}_i / Y_i) &= \beta_1 \\
\widetilde{Y}_i / Y_i &= e^{\beta_1} \\
(\widetilde{Y}_i - Y_i)/Y_i &= e^{\beta_1} - 1 \text{ From } X = 0 \text{ to } X = 1 \\
(\widetilde{Y}_i - Y_i)/Y_i &= e^{-\beta_1} - 1 \text{ From } X = 1 \text{ to } X = 0
\end{aligned}
$$

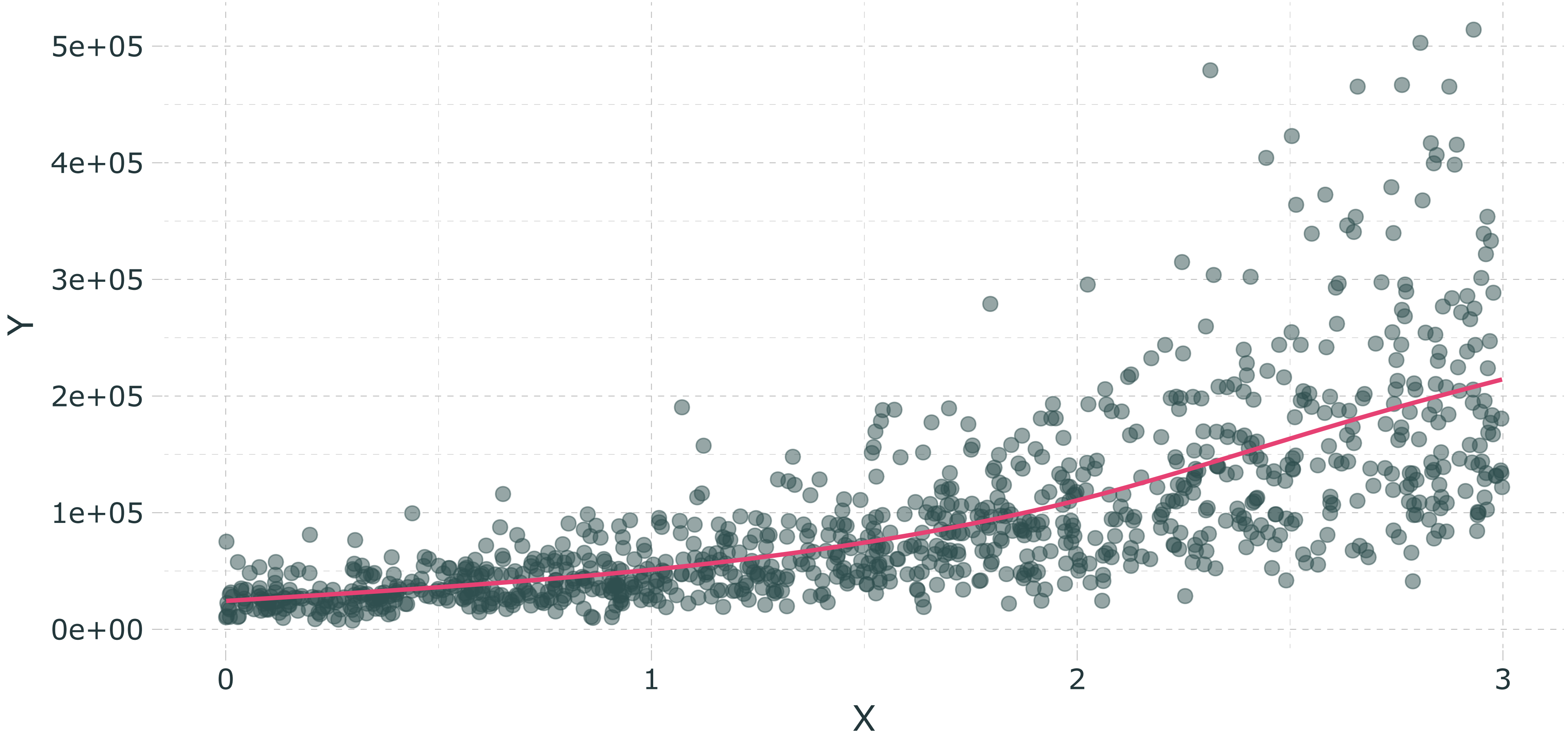
Approximation

$$
\begin{aligned}
\widetilde{\log}(Y_i) &= \log(Y_i) + \beta_1 \\
\widetilde{\log}(Y_i) &\approx \log(Y_i) + log(1 + \beta_1) \\
\widetilde{\log}(Y_i) &\approx \log(Y_i(1 + \beta_1)) \\
\widetilde{Y}_i &\approx Y_i(1 + \beta_1)
\end{aligned}
$$

- If we cannot re-scale `(X)` to have a small `(\beta)` we need to compute the percentage difference using the exact formula (left). Also, interpretation from 1 to 0 does not work well in approximation.

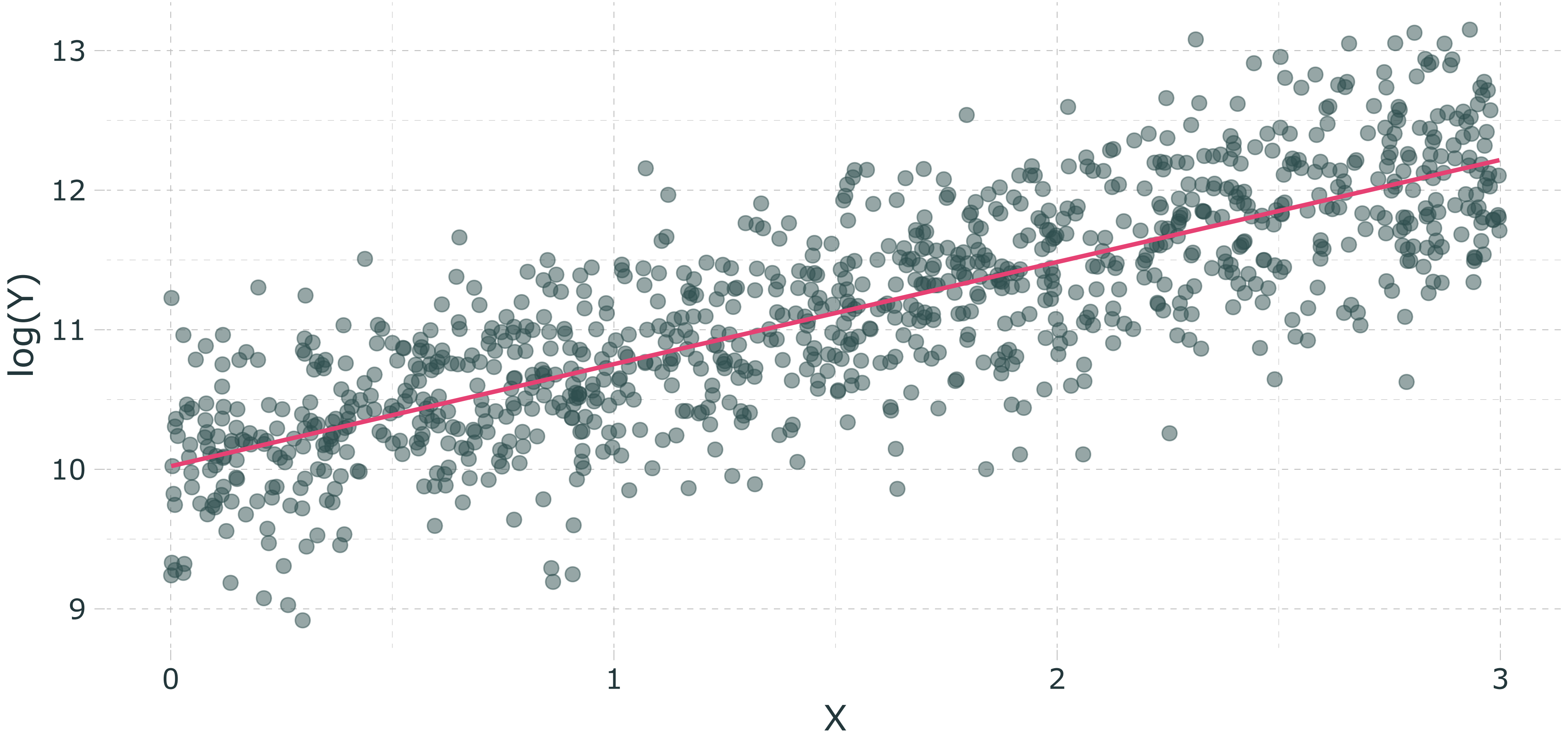
Log-Linear Example

$$\log(\hat{Y}_i) = 10.02 + 0.73 \cdot \mathrm{X}_i$$



Log-Linear Example

$$\log(\hat{Y}_i) = 10.02 + 0.73 \cdot \mathrm{X}_i$$



Log-Log Model

Nonlinear Model

$$Y_i = \alpha X_i^{\beta_1} e_i$$

- $Y > 0$, $X > 0$, and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Log-Log Model

Nonlinear Model

$$Y_i = \alpha X_i^{\beta_1} e_i$$

- $Y > 0$, $X > 0$, and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$\log(Y_i) = \log(\alpha) + \beta_1 \log(X_i) + \log(e_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(e_i) \equiv e_i$.

Log-Log Model

Nonlinear Model

$$Y_i = \alpha X_i^{\beta_1} e_i$$

- $Y > 0$, $X > 0$, and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$\log(Y_i) = \log(\alpha) + \beta_1 \log(X_i) + \log(e_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(e_i) \equiv e_i$.

Transformed (Linear) Model

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + e_i$$

Log-Log Model

Regression Model

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + e_i$$

Interpretation

- A one-percent increase in the explanatory variable leads to a β_1-percent change in the outcome variable, on average.
- This is the definition of an elasticity in economics ($\Delta\%Q/\Delta\%P$)
- *Example*: If $\log(\widehat{\text{Quantity Demanded}}_i) = 0.45 - 0.31 \cdot \log(\text{Income}_i)$, then each one-percent increase in income decreases quantity demanded by 0.31 percent.

Why?

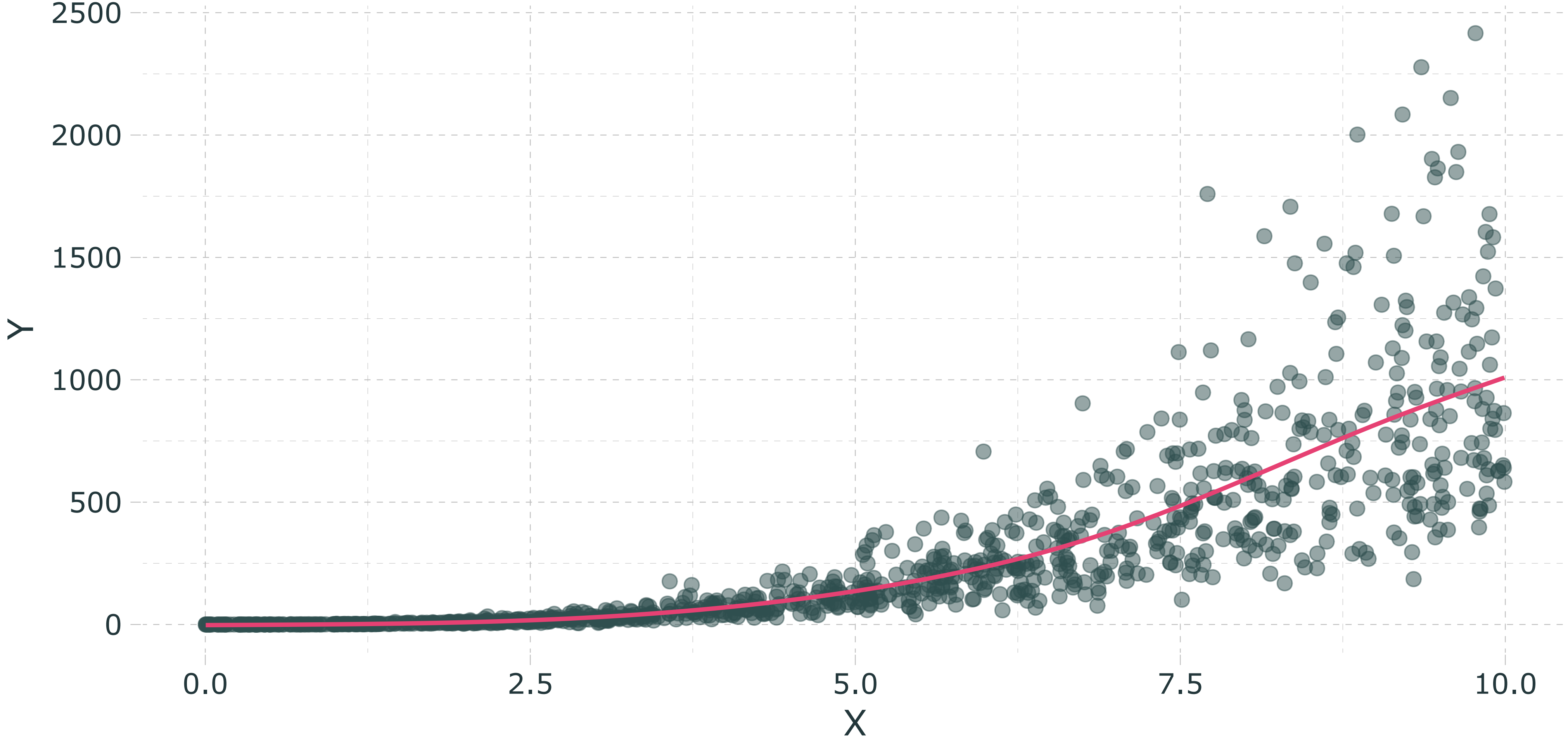
- We want to know how to interpret what is the associated increase in Y, when we increase X in **1 percent unit** (differnent from before).

$$
\begin{aligned}
\log(Y_i) &= \beta_0 + \beta_1 \log(X_i) + e_i \\
\widetilde{\log}(Y_i) &= \beta_0 + \beta_1 \log(X_i \times 1.01) + e_i \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_0 + \beta_1 X_i + \beta_1 \log(1.01) + e_i - \\
&\quad (\beta_0 + \beta_1 X_i + e_i) \\
\widetilde{\log}(Y_i) - \log(Y_i) &= \beta_1 \log(1.01) \\
\widetilde{\log}(Y_i) &= \log(Y_i) + \beta_1 \log(1.01) \\
\widetilde{\log}(Y_i) &\approx \log(Y_i) + \beta_1 \times 0.01 \\
\widetilde{\log}(Y_i) &\approx \log(Y_i) + \log(1 + \beta_1/100) \\
\widetilde{\log}(Y_i) &\approx \log(Y_i(1 + \beta_1/100))
\end{aligned}
$$

A one-percent increase in X leads to a β_1-percent increase in Y.

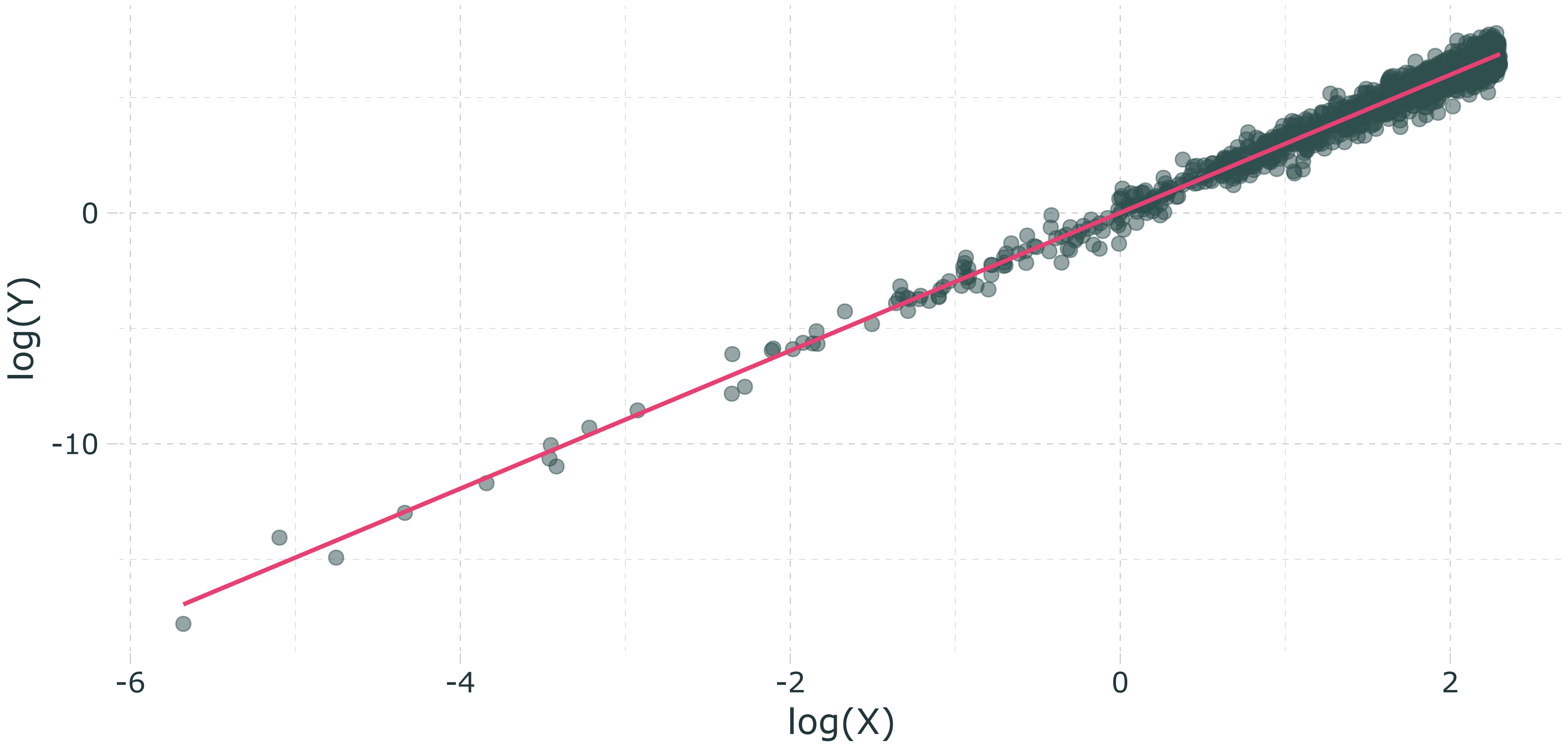
Log-Log Example

$$\log(\hat{Y}_i) = 0.01 + 2.99 \cdot \log(X_i)$$



Log-Log Example

$$\log(\hat{Y}_i) = 0.01 + 2.99 \cdot \log(\mathrm{X}_i)$$



Linear-Log Model

Nonlinear Model

$$e^{Y_i} = \alpha X_i^{\beta_1} e_i$$

- $X > 0$ and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Linear-Log Model

Nonlinear Model

$$e^{Y_i} = \alpha X_i^{\beta_1} e_i$$

- $X > 0$ and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$Y_i = \log(\alpha) + \beta_1 \log(X_i) + \log(e_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(e_i) \equiv e_i$.

Linear-Log Model

Nonlinear Model

$$e^{Y_i} = \alpha X_i^{\beta_1} e_i$$

- $X > 0$ and e_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$Y_i = \log(\alpha) + \beta_1 \log(X_i) + \log(e_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(e_i) \equiv e_i$.

Transformed (Linear) Model

$$Y_i = \beta_0 + \beta_1 \log(X_i) + e_i$$

Linear-Log Model

Regression Model

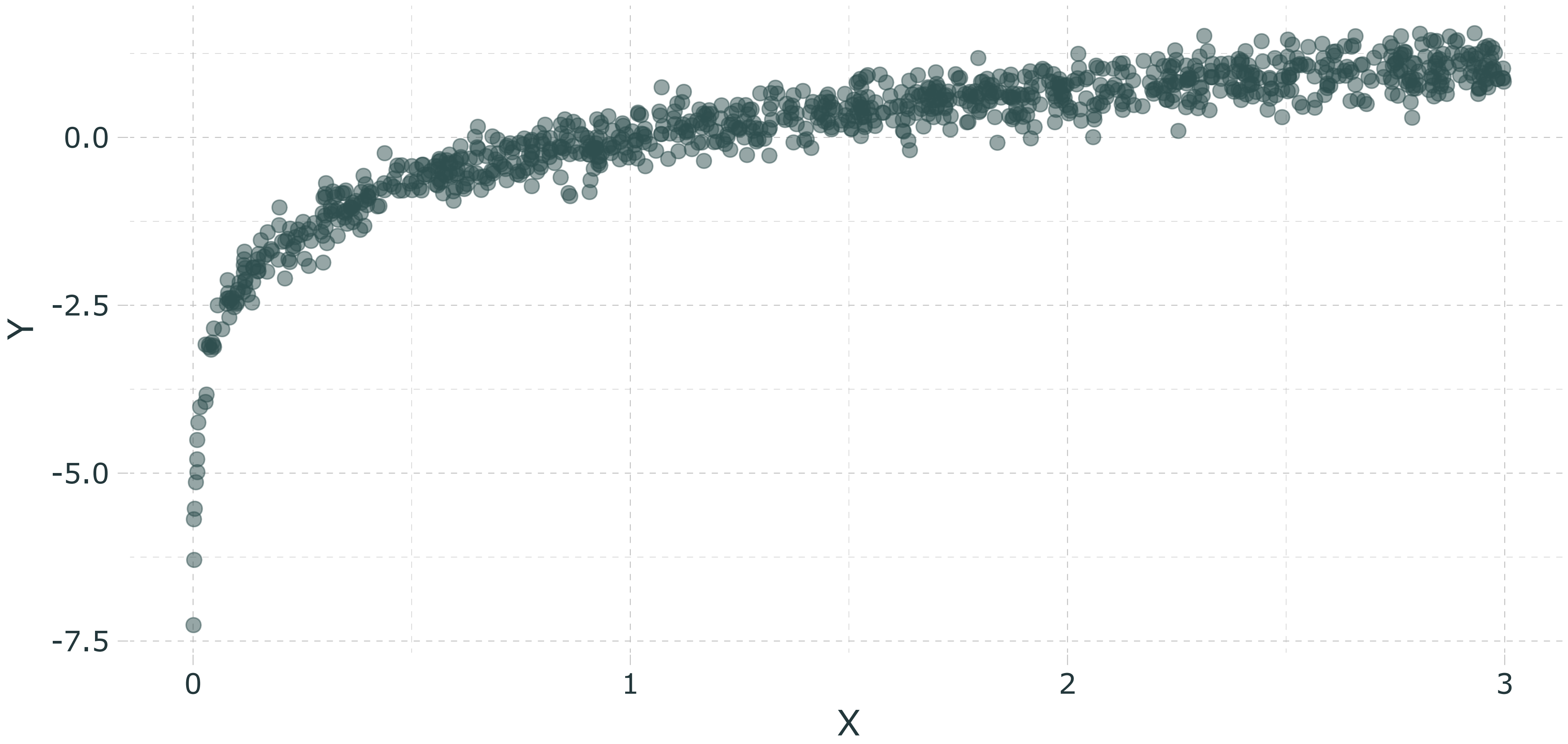
$$Y_i = \beta_0 + \beta_1 \log(X_i) + e_i$$

Interpretation

- A one-percent increase in the explanatory variable increases the outcome variable by approximately $\beta_1 \div 100$, on average.
- *Example:* If $(\hat{\text{Blood Pressure}})_i = 150 - 9.1 \log(\text{Income}_i)$, then a one-percent increase in income decrease blood pressure by 0.091 points.

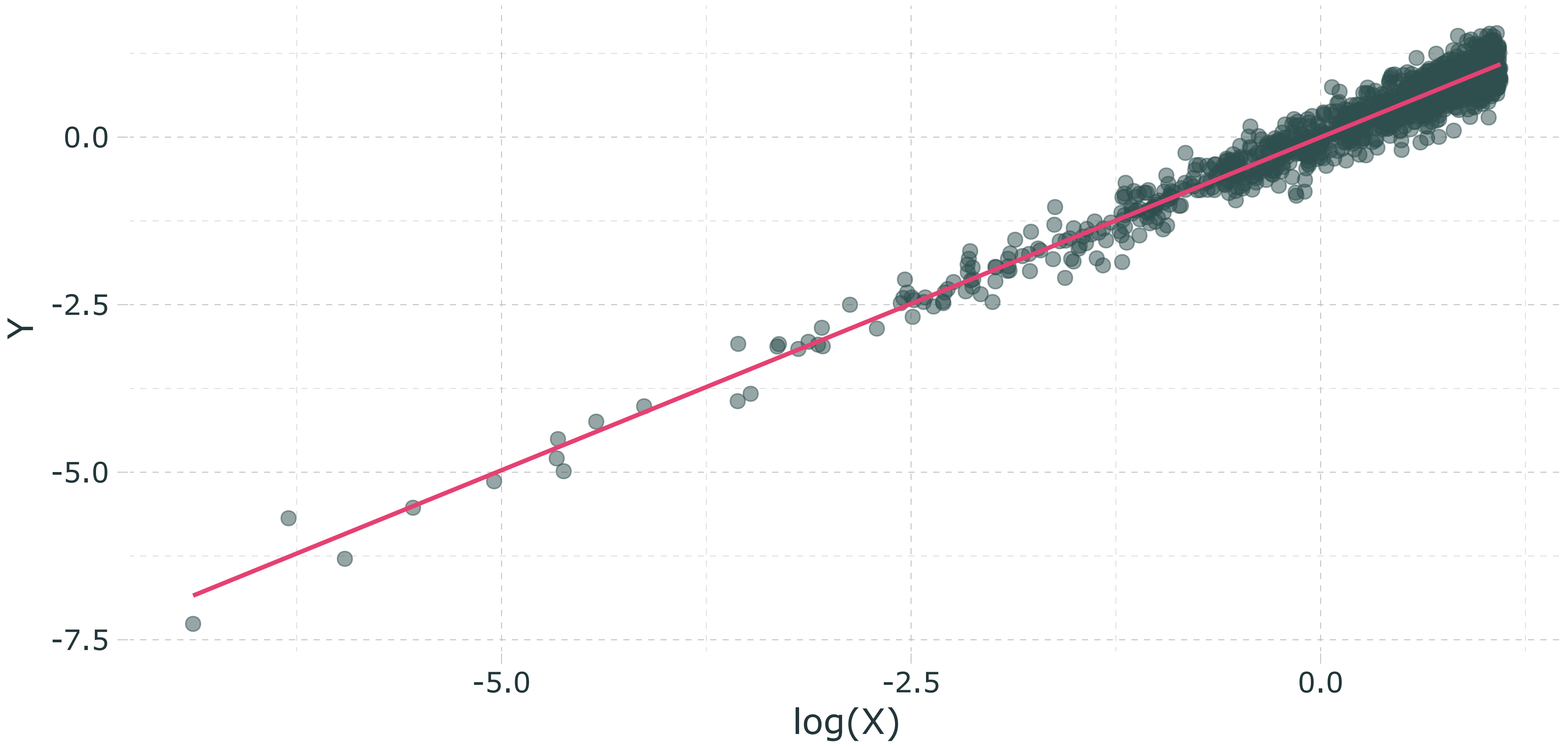
Linear-Log Example

$$\hat{Y}_i = 0 + 0.99 \cdot \log(X_i)$$



Linear-Log Example

$$\hat{Y}_i = 0 + 0.99 \cdot \log(X_i)$$

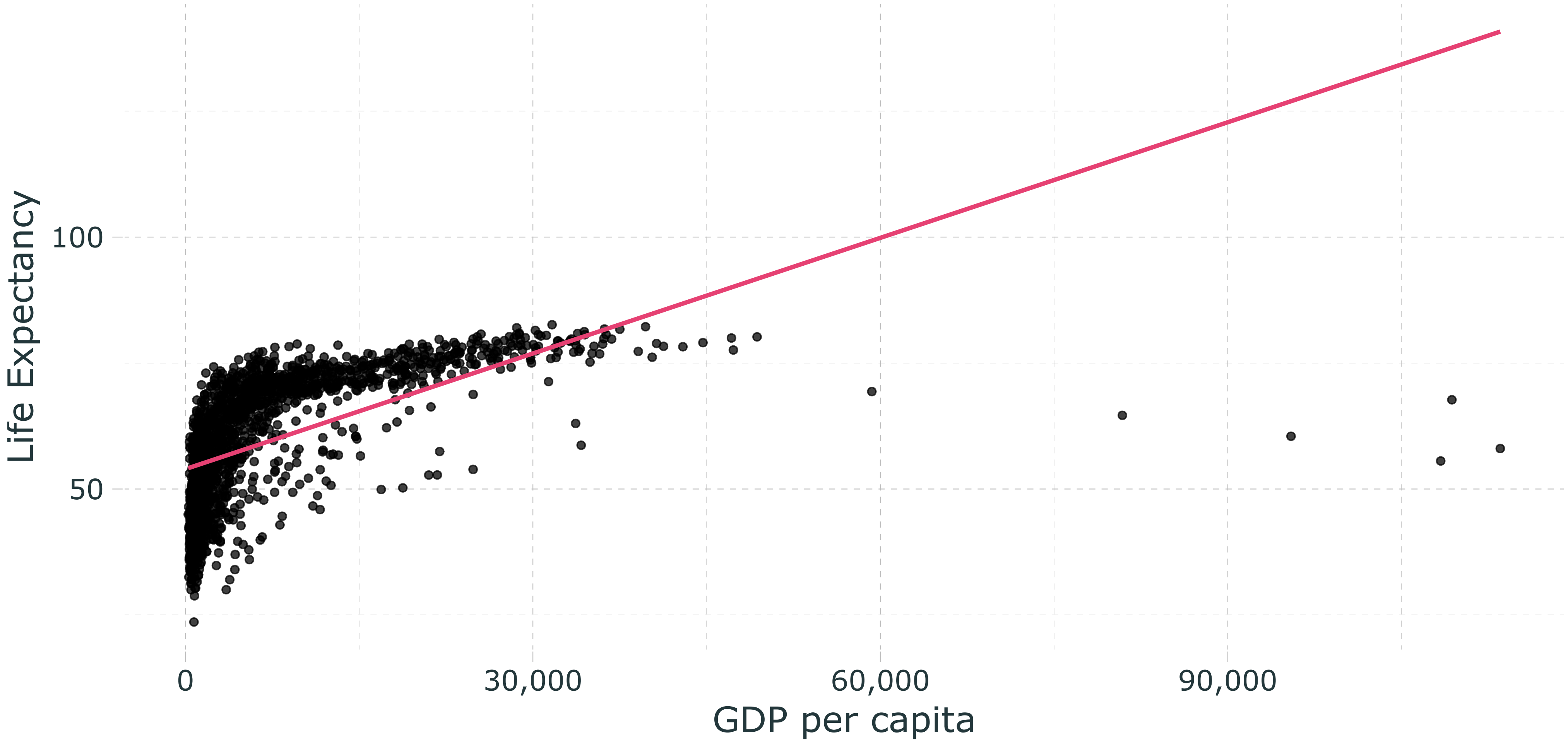


(Approximate) Coefficient Interpretation

Model	β_1 **Interpretation**
Level-level $Y_i = \beta_0 + \beta_1 X_i + e_i$	$\Delta Y = \beta_1 \cdot \Delta X$ *A one-unit increase in X leads to a β_1-unit increase in Y*
Log-level $\log(Y_i) = \beta_0 + \beta_1 X_i + e_i$	$\%\Delta Y = 100 \cdot \beta_1 \cdot \Delta X$ *A one-unit increase in X leads to a $\beta_1 \cdot 100\%$ increase in Y*
Log-log $\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + e_i$	$\%\Delta Y = \beta_1 \cdot \%\Delta X$ *A one-percent increase in X leads to a $\beta_1\%$ increase in Y*
Level-log $Y_i = \beta_0 + \beta_1 \log(X_i) + e_i$	$\Delta Y = (\beta_1 \div 100) \cdot \%\Delta X$ *A one-percent increase in X leads to a $\beta_1 \div 100$-unit increase in Y*

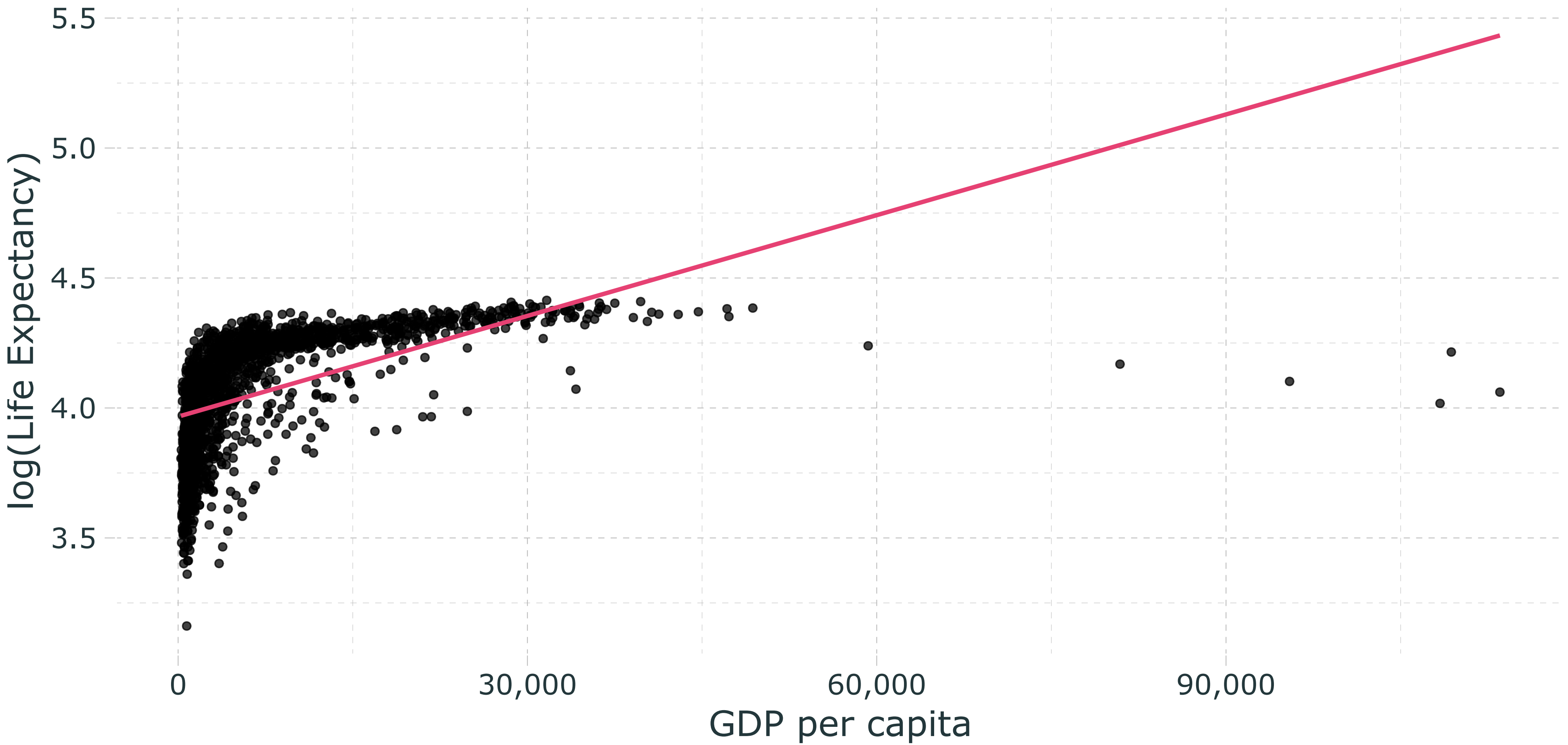
Can We Do Better?

$$\widehat{(\text{Life Expectancy})_i} = 53.96 + 8 \times 10^{-4} \cdot \text{GDP}_i \qquad R^2 = 0.34$$



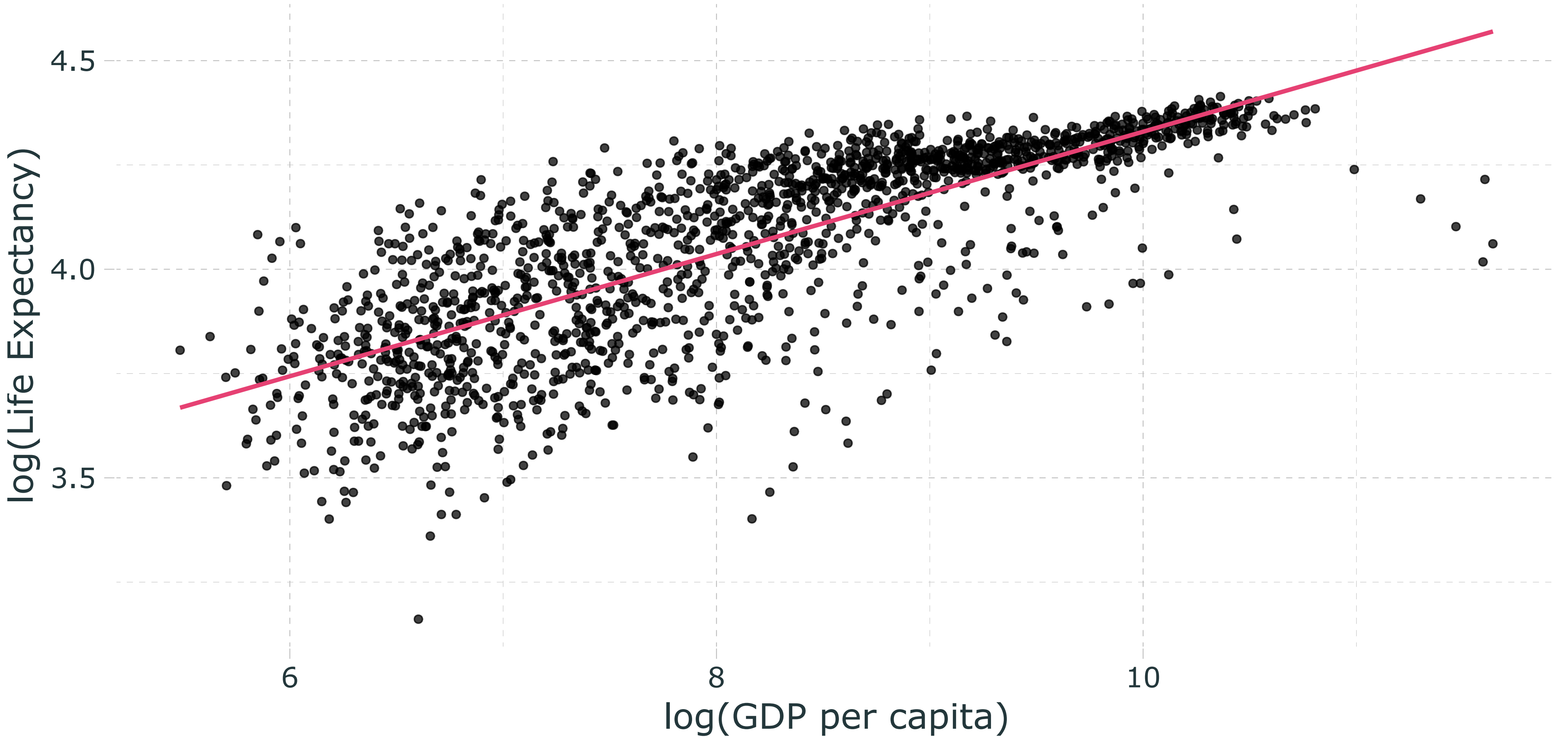
Can We Do Better?

$$\log(\widehat{\text{Life Expectancy}_i}) = 3.97 + 1.3 \times 10^{-5} \cdot \text{GDP}_i \qquad R^2 = 0.3$$



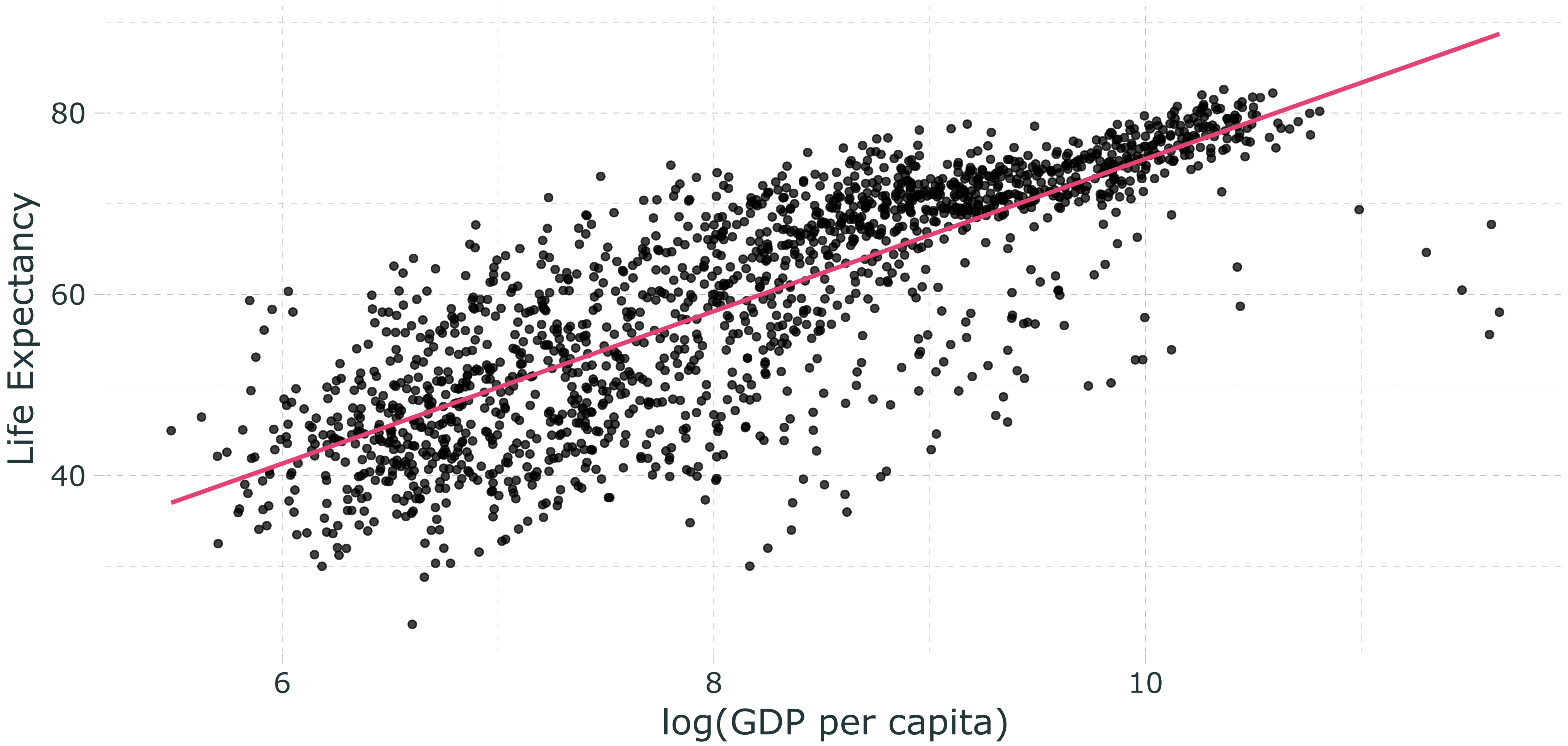
Can We Do Better?

$$\widehat{\log(\text{Life Expectancy}_i)} = 2.86 + 0.15 \cdot \log(\text{GDP}_i) \qquad R^2 = 0.61$$



Can We Do Better?

$$\widehat{(\text{Life Expectancy})}_i = -9.1 + 8.41 \cdot \log(\text{GDP}_i) \qquad R^2 = 0.65$$



Practical Considerations

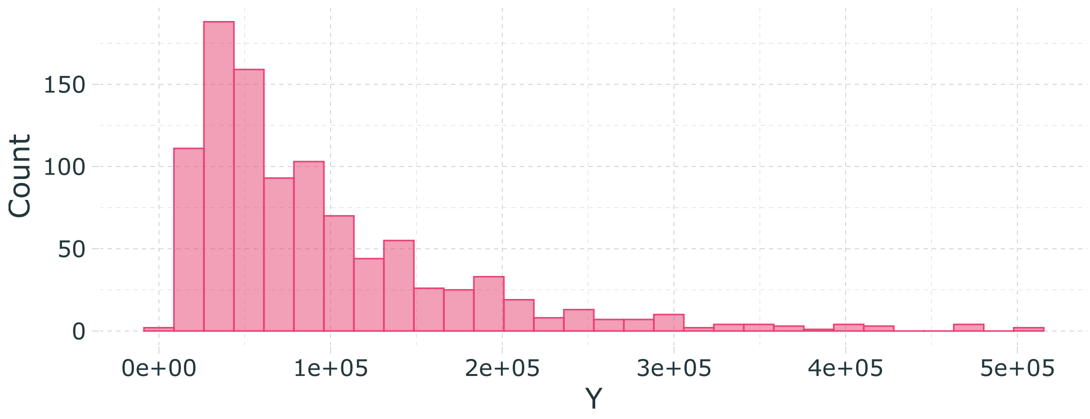
Consideration 1: Do your data take negative numbers or zeros as values?

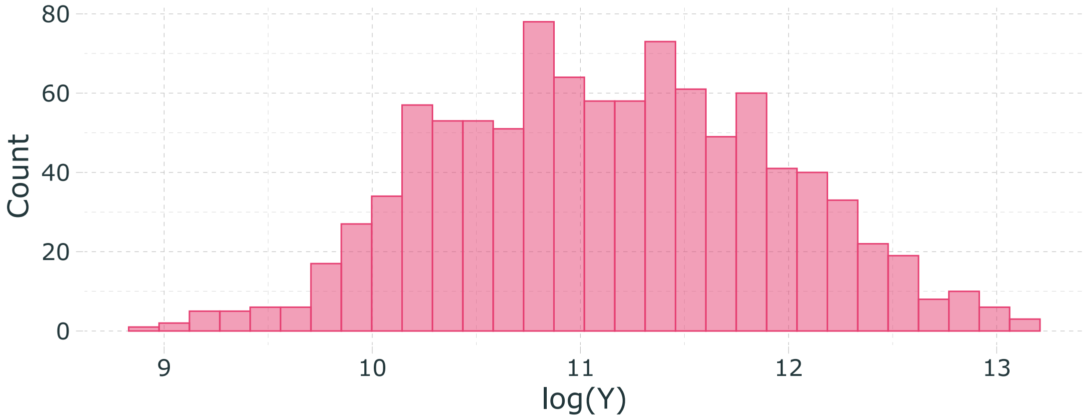
```
log(0)
```

```
#> [1] -Inf
```

Consideration 2: What coefficient interpretation do you want? Unit change? Unit-free percent change?

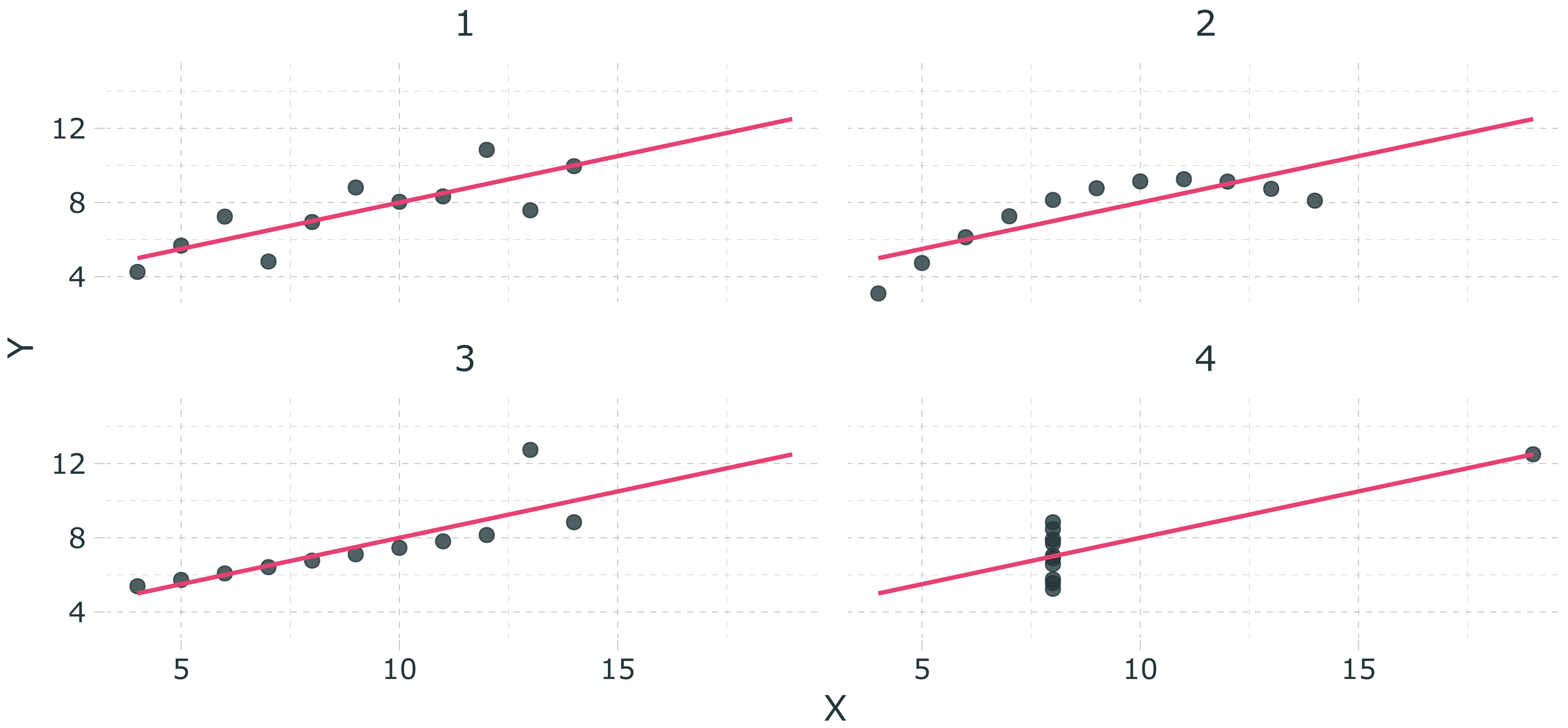
Consideration 3: Are your data skewed?





Final Message: Allways Plot Your Data (Anscombe's Quartet)

Four "*identical*" regressions: Intercept = 3, Slope = 0.5, R^2 = 0.67



Other Non-linear Relationships

- Binary dependent variable
- Interactions (covered later in the course)
- Polynomial regressors (not covered)

Binary Dependent Variable

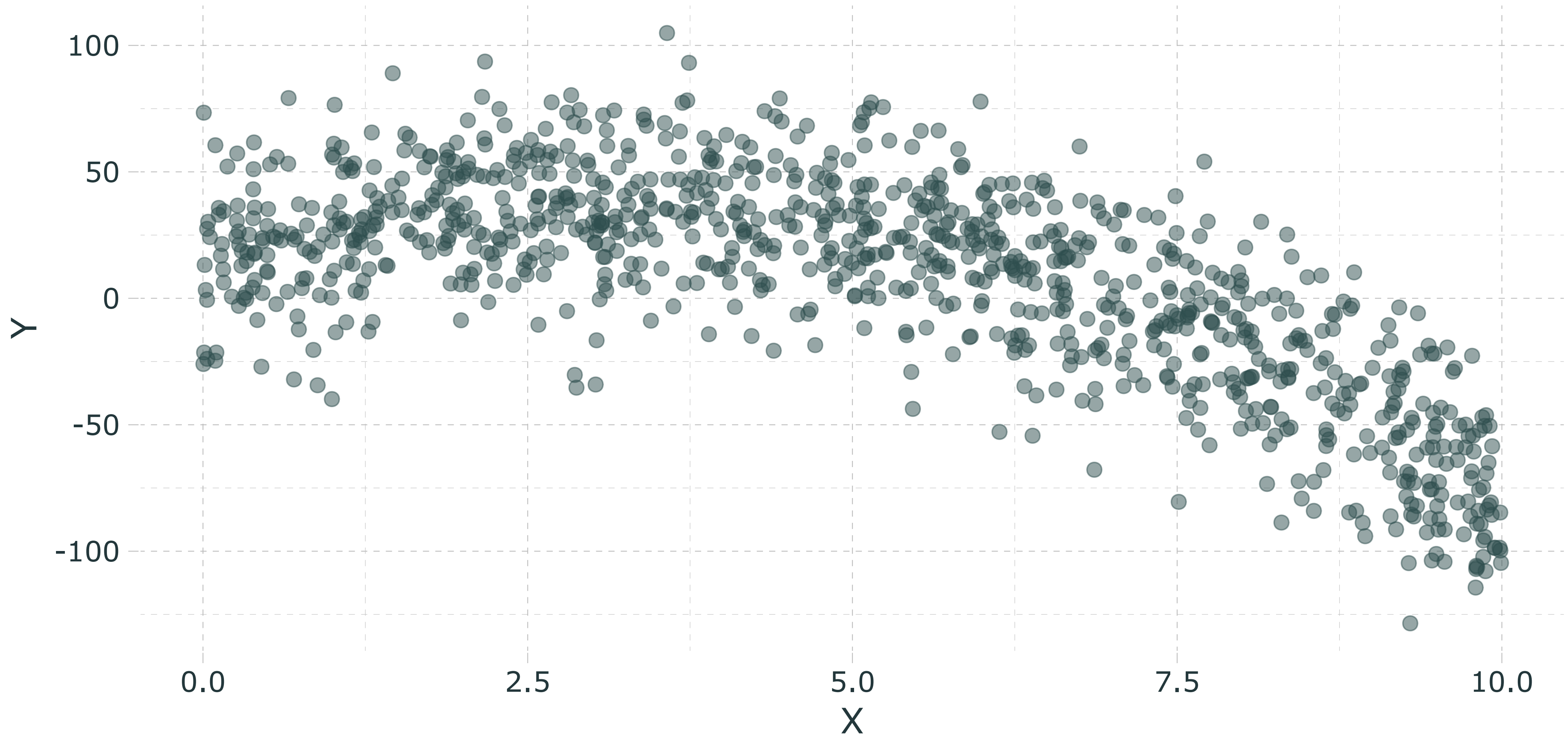
- Previously, introductory courses spent significant time arguing that binary dependent outcomes invalidated regression.
- The two main reasons were:
  1. This is a highly non-linear relationship (draw plot)
  2. The errors in this context have a variance that is correlated with the Xs (heteroskedasticity).
- The approach we follow here does not focus on spending much time addressing this concerns. Because
  - Even if its non-linear, the CEF property #2 says that regression will find the best linear approximation. The key is to choose regressors well (in this case a collection of dummies probably will work better than a single slope).
  - We now use robust standard errors pretty much all the time.
- Regression in this context takes the name Linear Probability Model (the other methods not covered here are Logit and Probit estimation).

(Polynomials Terms in a Regression)

(Not Covered, but leaving it here in case you are interested.

Requires a little knowing the derivative of polynomials)
(Will mark each of the non-covered slides with an [NC])

Quadratic (and other Polynomial) Relationships [NC]



Quadratic Regression [NC]

Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

Interpretation

Sign of β_2 indicates whether the relationship is convex (+) or concave (-)

Sign of β_1? 🤷‍♂️

Partial derivative of Y with respect to X is the **marginal effect** of X on Y:

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$$

- Effect of X depends on the level of X

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26       5.81 8.30e-  9
#> 2 x              15.7      1.03      15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982   -25.4  2.46e-110
```

What is the marginal effect of X on Y?

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2     2.26        5.81 8.30e-  9
#> 2 x              15.7     1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50    0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y?

$$\widehat{\frac{\partial \text{Y}}{\partial \text{X}}} = \hat{\beta}_1 + 2\hat{\beta}_2 X = 15.69 + -4.99X$$

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2     2.26        5.81 8.30e-  9
#> 2 x              15.7     1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50    0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 0$?

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26       5.81 8.30e-  9
#> 2 x              15.7      1.03      15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982   -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 0$?

$$\left.\widehat{\frac{\partial \mathrm{Y}}{\partial \mathrm{X}}}\right|_{\mathrm{X}=0} = \hat{\beta}_1 = 15.69$$

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2     2.26        5.81 8.30e-  9
#> 2 x              15.7     1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50    0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 2$?

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26        5.81 8.30e-  9
#> 2 x              15.7      1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 2$?

$$\widehat{\frac{\partial \mathrm{Y}}{\partial \mathrm{X}}}\bigg|_{\mathrm{X}=2} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot (2) = 15.69 - 9.99 = 5.71$$

Quadratic Regression [NC]

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26       5.81 8.30e-  9
#> 2 x              15.7      1.03      15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982   -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 7$?

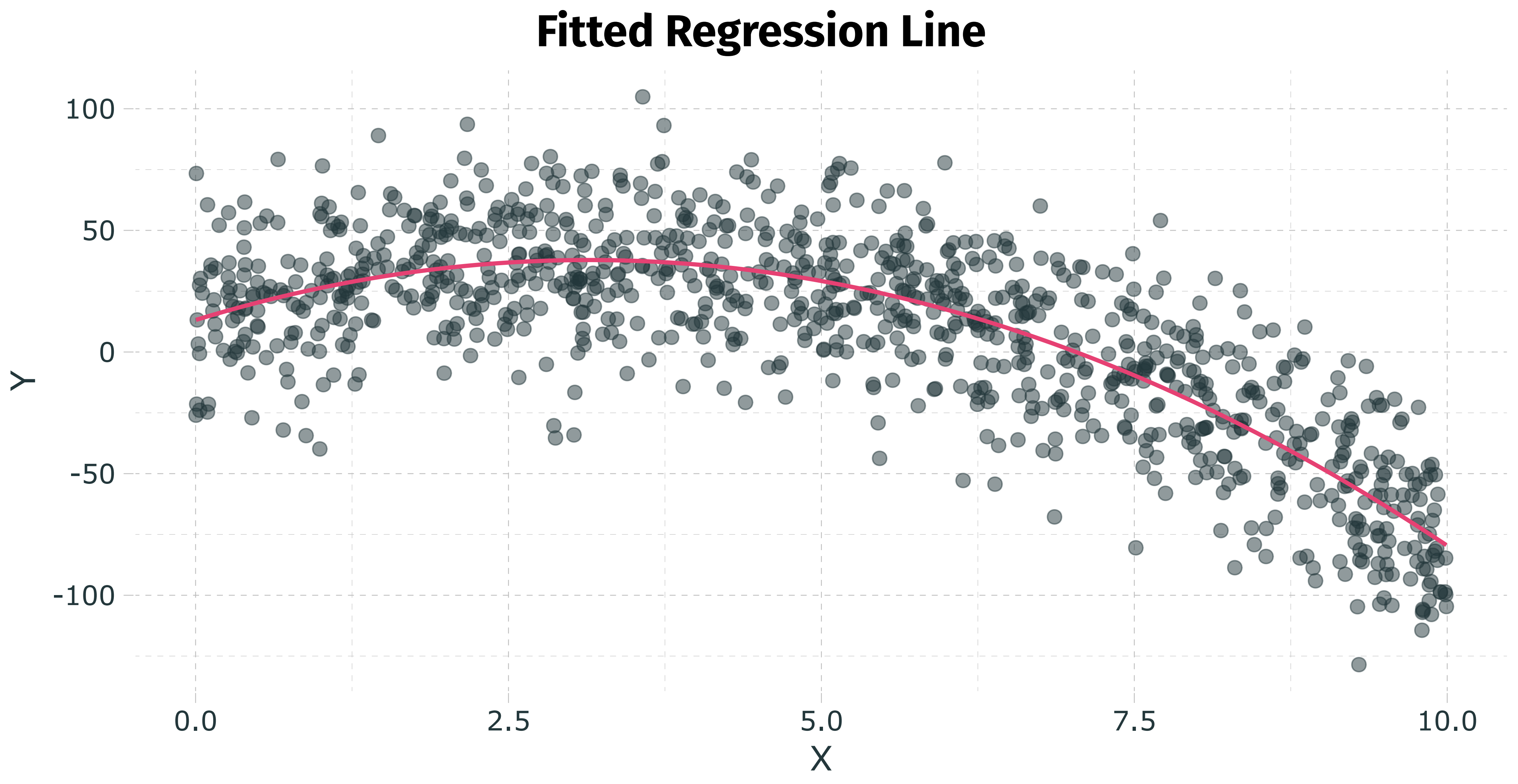
Quadratic Regression [NC]

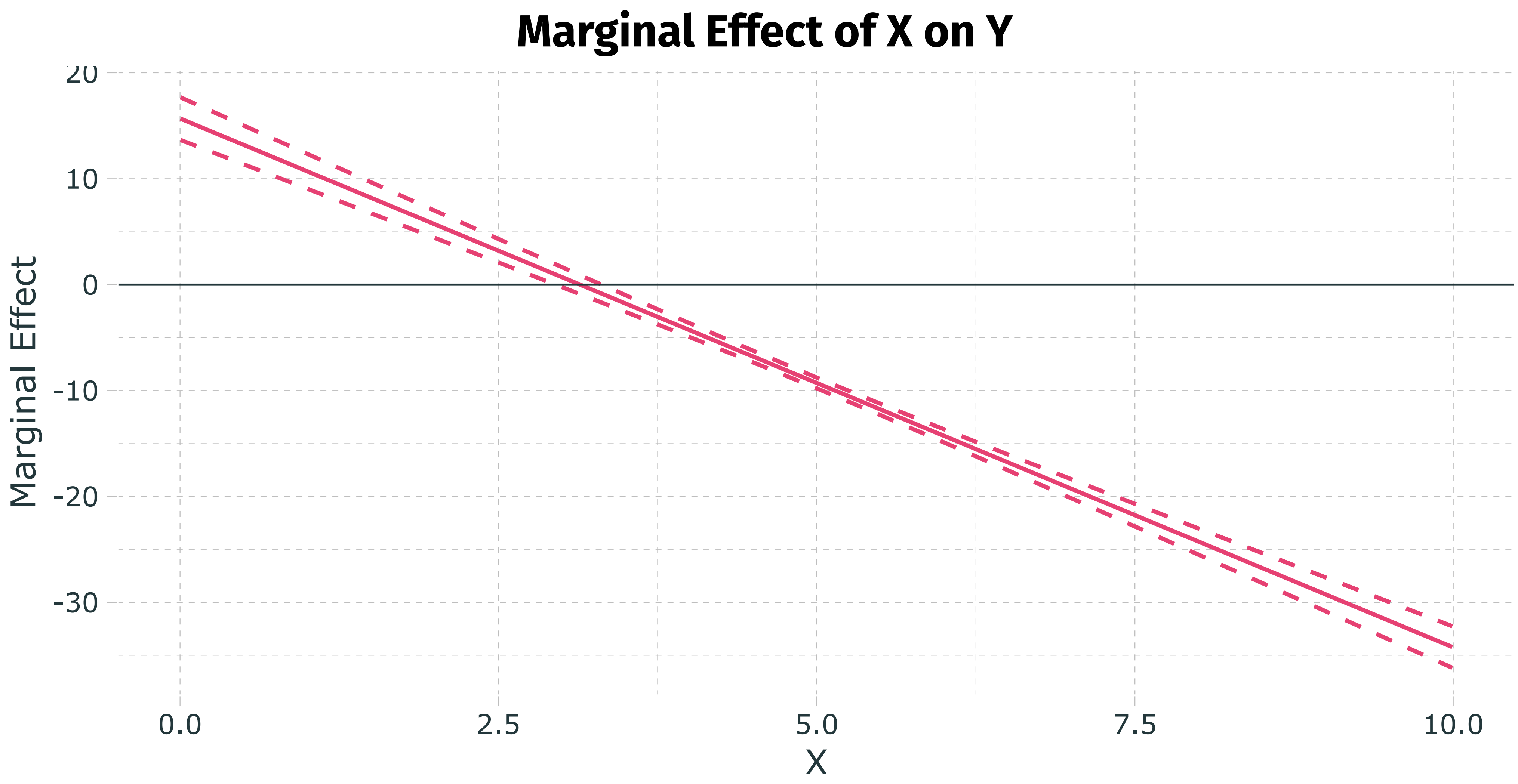
```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 × 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26       5.81 8.30e-  9
#> 2 x              15.7      1.03      15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982   -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 7$?

$$\widehat{\frac{\partial Y}{\partial X}}\Bigg|_{X=7} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot (7) = 15.69 - 34.96 = -19.27$$


Fitted Regression Line
100
50
0
-50
-100
Y
0.0
2.5
5.0
7.5
10.0
X



Marginal Effect of X on Y
Marginal Effect
20
10
0
-10
-20
-30
0.0
2.5
5.0
7.5
10.0
X

Quadratic Regression [NC]

Where does the regression $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$ *turn*?

- In other words, where is the peak (valley) of the fitted relationship?

Step 1: Take the derivative and set equal to zero.

$$\frac{\widehat{\partial \mathrm{Y}}}{\partial \mathrm{X}} = \hat{\beta}_1 + 2\hat{\beta}_2 X = 0$$

Step 2: Solve for X.

$$X = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

Example: Peak of previous regression occurs at $X = 3.14$.

Acknowledgments

- Kyle Raze's Undergraduate Econometrics 1
- MM
- Science Po Econometrics Course
- Nick Huntington-Klein's Explanation of logs