Nonlinear Regression Functions

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.
Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Contents

In this lesson you will learn ...

- modeling nonlinear regression functions
- polynomial, exponentials, logarithms
- interactions between independent variables
- interactions involving continuous and binary variables

In this lesson you will learn …

- modeling nonlinear regression functions
- polynomial, exponentials, logarithms
- interactions between independent variables
- interactions involving continuous and binary variables

In this lesson you will learn …

- modeling nonlinear regression functions
- polynomial, exponentials, logarithms
- interactions between independent variables
- interactions involving continuous and binary variables

In this lesson you will learn …

- modeling nonlinear regression functions
- polynomial, exponentials, logarithms
- interactions between independent variables
- interactions involving continuous and binary variables

Nonlinear Regression

Modeling Nonlinear Regression Functions

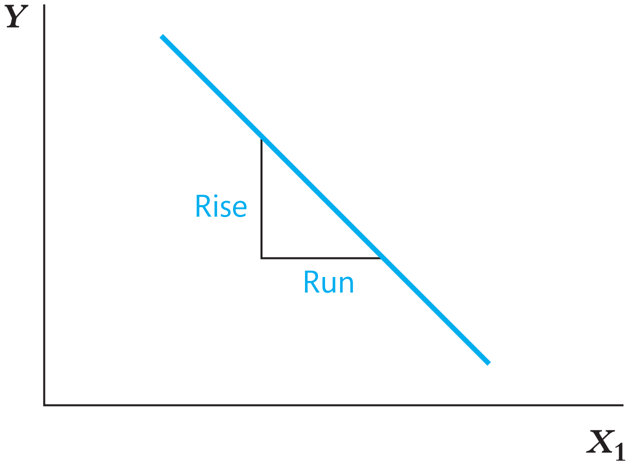
- **Constant slope:**
  The effect on Y of a unit change in X is the same for all values of the regressors.
- In the linear regression model, the population regression function has a constant slope.
- **Non-constant slope:**
  If the effect on Y of a change in X in fact depends on the value of one or more of the regressors, the population regression function is nonlinear.

- **Constant slope:**
  The effect on Y of a unit change in X is the same for all values of the regressors.
- In the linear regression model, the population regression function has a constant slope.
- **Non-constant slope:**
  If the effect on Y of a change in X in fact depends on the value of one or more of the regressors, the population regression function is nonlinear.

Modeling Nonlinear Regression Functions

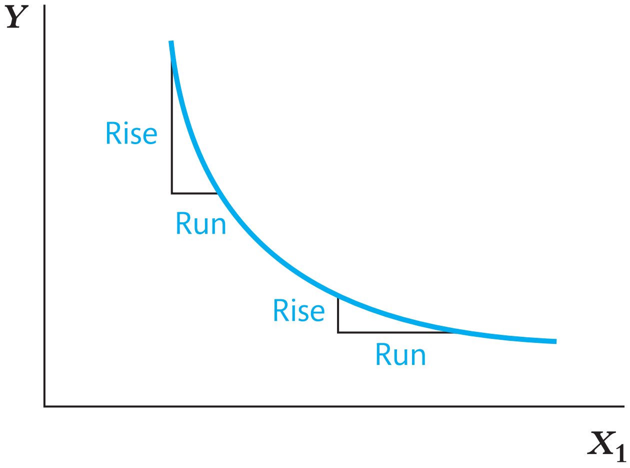
- **Constant slope:**
  The effect on Y of a unit change in X is the same for all values of the regressors.
- In the linear regression model, the population regression function has a constant slope.
- **Non-constant slope:**
  If the effect on Y of a change in X in fact depends on the value of one or more of the regressors, the population regression function is nonlinear.

Population Regression Functions with Different Slopes



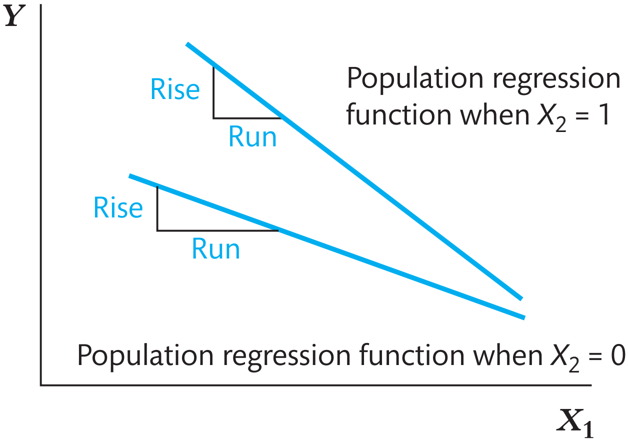
(a) Constant slope

Population Regression Functions with Different Slopes



(b) Slope depends on the value of X_1

Population Regression Functions with Different Slopes



(c) Slope depends on the value of X_2

Modeling Nonlinearities

Nonlinear Functions of a Single Independent Variable

Modeling Nonlinearities Using Multiple Regression:

1. Identify a possible nonlinear relationship.
2. Specify a nonlinear function, and estimate its parameters by OLS.
3. Determine whether the nonlinear model improves upon a linear model.
4. Plot the estimated nonlinear regression function.
5. Estimate the effect on Y of a change in X.

Nonlinear Functions of a Single Independent Variable

Modeling Nonlinearities Using Multiple Regression:

1. Identify a possible nonlinear relationship.
2. Specify a nonlinear function, and estimate its parameters by OLS.
3. Determine whether the nonlinear model improves upon a linear model.
4. Plot the estimated nonlinear regression function.
5. Estimate the effect on Y of a change in X.

Modeling Nonlinearities Using Multiple Regression:

1. Identify a possible nonlinear relationship.
2. Specify a nonlinear function, and estimate its parameters by OLS.
3. Determine whether the nonlinear model improves upon a linear model.
4. Plot the estimated nonlinear regression function.
5. Estimate the effect on Y of a change in X.

Modeling Nonlinearities Using Multiple Regression:

1. Identify a possible nonlinear relationship.
2. Specify a nonlinear function, and estimate its parameters by OLS.
3. Determine whether the nonlinear model improves upon a linear model.
4. Plot the estimated nonlinear regression function.
5. Estimate the effect on Y of a change in X.

Nonlinear Functions of a Single Independent Variable

Modeling Nonlinearities Using Multiple Regression:

1. Identify a possible nonlinear relationship.
2. Specify a nonlinear function, and estimate its parameters by OLS.
3. Determine whether the nonlinear model improves upon a linear model.
4. Plot the estimated nonlinear regression function.
5. Estimate the effect on Y of a change in X.

Quadratic Effects On Test Scores

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

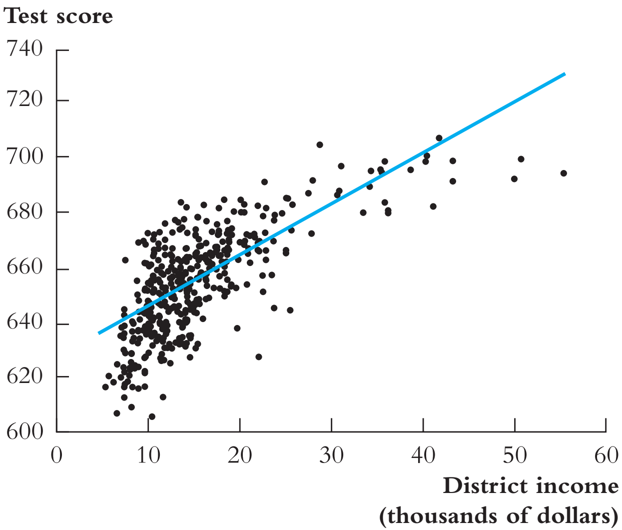
- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Nonlinear Effects on Test Scores of the Student-Teacher Ratio

California Test Score Data:

- The economic background of the students is an important factor in explaining performance on standardized tests.
- Economic background variables: Measure the fraction of students in the district who come from poor families.
  - The percentage of students qualifying for a subsidized lunch.
  - The percentage of students whose families qualify for income assistance.
  - Average annual per capita income in the school district.
- Students from affluent districts do better on the tests than students from poor districts.
- Test scores and district income are strongly positively correlated, with Pearson's correlation coefficient $r = 0.71$.
- Some curvature in the relationship between test scores and district income is not captured by the linear regression.

Linear Regression Function



Test Scores and District Income

Quadratic regression model:

- A quadratic population regression model relating test scores and income:

$$TestScore_i = \beta_0 + \beta_1\, Income_i + \beta_2\, (Income_i)^2 + u_i$$

- Estimate the quadratic equation by OLS:

$$\underset{}{\widehat{TestScore}} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85}\ Income - \underset{(0.0048)}{0.0423}\ Income^2 \quad \bar{R}^2 = 0.554$$

- The quadratic function captures the curvature in the scatterplot: It is steep for low values of district income but flattens out when district income is high.
- Test $H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_2 - 0}{\text{SE}(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis.

Test Scores and District Income

Quadratic regression model:

- A quadratic population regression model relating test scores and income:

$$TestScore_i = \beta_0 + \beta_1 \, Income_i + \beta_2 \, (Income_i)^2 + u_i$$

- Estimate the quadratic equation by OLS:

$$\widehat{TestScore} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85} \; Income - \underset{(0.0048)}{0.0423} \; Income^2 \quad \bar{R}^2 = 0.554$$

- The quadratic function captures the curvature in the scatterplot: It is steep for low values of district income but flattens out when district income is high.
- Test $H_0: \; \beta_2 = 0, \quad H_1: \; \beta_2 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_2 - 0}{\text{SE}(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis.

Test Scores and District Income

Quadratic regression model:

- A quadratic population regression model relating test scores and income:

$$TestScore_i = \beta_0 + \beta_1 \, Income_i + \beta_2 \, (Income_i)^2 + u_i$$

- Estimate the quadratic equation by OLS:

$$\underset{}{\widehat{TestScore}} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85} \; Income - \underset{(0.0048)}{0.0423} \; Income^2 \quad \bar{R}^2 = 0.554$$

- The quadratic function captures the curvature in the scatterplot: It is steep for low values of district income but flattens out when district income is high.
- Test $H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_2 - 0}{\text{SE}(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis.

Test Scores and District Income

Quadratic regression model:

- A quadratic population regression model relating test scores and income:

$$TestScore_i = \beta_0 + \beta_1\, Income_i + \beta_2\, (Income_i)^2 + u_i$$

- Estimate the quadratic equation by OLS:

$$\begin{array}{lllll} \widehat{TestScore} = 607.3 + & 3.85 & Income - & 0.0423 & Income^2 \quad \bar{R}^2 = 0.554 \\ \qquad\qquad\quad (2.9) & (0.27) & & (0.0048) & \end{array}$$

- The quadratic function captures the curvature in the scatterplot: It is steep for low values of district income but flattens out when district income is high.
- Test $H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_2 - 0}{\text{SE}(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis.

Test Scores and District Income

Quadratic regression model:

- A quadratic population regression model relating test scores and income:

$$TestScore_i = \beta_0 + \beta_1\, Income_i + \beta_2\, (Income_i)^2 + u_i$$

- Estimate the quadratic equation by OLS:

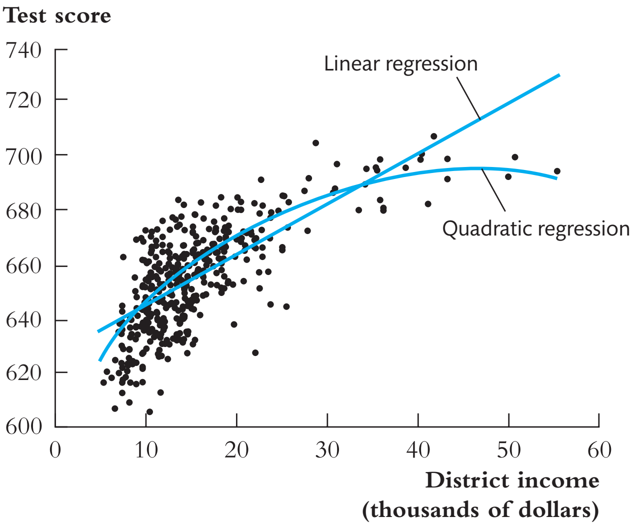
$$\begin{array}{lllll} \widehat{TestScore} = 607.3 + & 3.85 & Income - & 0.0423 & Income^2 \quad \bar{R}^2 = 0.554 \\ \qquad\qquad\quad (2.9) & (0.27) & & (0.0048) & \end{array}$$

- The quadratic function captures the curvature in the scatterplot: It is steep for low values of district income but flattens out when district income is high.
- Test $H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_2 - 0}{\text{SE}(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis.

Quadratic Regression Function



Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:
  - An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.
  - An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:
  - An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.
  - An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:

Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:
  - An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.
  - An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:
  - An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.
  - An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

Test Scores and District Income

- What is the predicted change in test scores associated with a change in district income of $\$1000$, based on the estimated quadratic regression function?
- In the linear regression, the regression coefficients had a natural interpretation — Not so in a non-linear regression.
- The effect depends on the initial district income: The slope of the estimated quadratic regression function is steeper at low values of income.
- Consider two cases:
  - An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.
  - An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

Test Scores and District Income

Predicted change in test scores:

- An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.

$$\begin{aligned}\Delta\hat{Y} &= (\hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2) \\ &= (607.3 + 3.85 \times 11 - 0.0423 \times 11^2) - (607.3 + 3.85 \times 10 - 0.0423 \times 10^2) \\ &= 644.53 - 641.57 \\ &= 2.96\end{aligned}$$

- An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

$$\begin{aligned}\Delta\hat{Y} &= (\hat{\beta}_0 + \hat{\beta}_1 \times 41 + \hat{\beta}_2 \times 41^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 40 + \hat{\beta}_2 \times 40^2) \\ &= (607.3 + 3.85 \times 41 - 0.0423 \times 41^2) - (607.3 + 3.85 \times 40 - 0.0423 \times 40^2) \\ &= 694.04 - 693.62 \\ &= 0.42\end{aligned}$$

Test Scores and District Income

Predicted change in test scores:

- An increase in district income from $\$10,000$ per capita to $\$11,000$ per capita.

$$
\begin{aligned}
\Delta\hat{Y} &= (\hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2) \\
&= (607.3 + 3.85 \times 11 - 0.0423 \times 11^2) - (607.3 + 3.85 \times 10 - 0.0423 \times 10^2) \\
&= 644.53 - 641.57 \\
&= 2.96
\end{aligned}
$$

- An increase in district income from $\$40,000$ per capita to $\$41,000$ per capita.

$$
\begin{aligned}
\Delta\hat{Y} &= (\hat{\beta}_0 + \hat{\beta}_1 \times 41 + \hat{\beta}_2 \times 41^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 40 + \hat{\beta}_2 \times 40^2) \\
&= (607.3 + 3.85 \times 41 - 0.0423 \times 41^2) - (607.3 + 3.85 \times 40 - 0.0423 \times 40^2) \\
&= 694.04 - 693.62 \\
&= 0.42
\end{aligned}
$$

Test Scores and District Income

Standard errors of estimated effects:

- Linear regression:

$$\text{SE}(\Delta\hat{Y}) = \text{SE}(\hat{\beta}_1)\Delta X_1$$
$$\hat{\beta}_1\Delta X_1 \pm 1.96\,\text{SE}(\hat{\beta}_1)\Delta X_1$$

- Non-Linear regression:

$$\Delta\hat{Y} = \hat{\beta}_1 \times (11-10) + \hat{\beta}_2 \times (11^2 - 10^2) = \hat{\beta}_1 + 21\hat{\beta}_2$$
$$\text{SE}(\Delta\hat{Y}) = \text{SE}(\hat{\beta}_1 + 21\hat{\beta}_2)$$

Test Scores and District Income

Standard errors of estimated effects:

- Linear regression:

$$\mathsf{SE}(\Delta\hat{Y}) = \mathsf{SE}(\hat{\beta}_1)\Delta X_1$$

$$\hat{\beta}_1 \Delta X_1 \pm 1.96\, \mathsf{SE}(\hat{\beta}_1)\Delta X_1$$

- Non-Linear regression:

$$\Delta\hat{Y} = \hat{\beta}_1 \times (11 - 10) + \hat{\beta}_2 \times (11^2 - 10^2) = \hat{\beta}_1 + 21\hat{\beta}_2$$

$$\mathsf{SE}(\Delta\hat{Y}) = \mathsf{SE}(\hat{\beta}_1 + 21\hat{\beta}_2)$$

Test Scores and District Income

Standard errors of estimated effects:

- To compute the standard error of $\hat{\beta}_1 + 21\hat{\beta}_2$, consider the F-statistic of

$$H_0: \beta_1 + 21\beta_2 = 0, \quad H_1: \beta_1 + 21\beta_2 \neq 0$$

Since $F = 299.94$ and $\Delta\hat{Y} = 2.96$,

$$\text{SE}(\Delta\hat{Y}) = \frac{|\Delta\hat{Y}|}{\sqrt{F}} = \frac{2.96}{\sqrt{299.94}} \approx 0.17$$

- A 95% confidence interval for the change in the expected value of Y is:

$$2.96 \pm 1.96 \times 0.17 = (2.63, 3.29)$$

Test Scores and District Income

Standard errors of estimated effects:

- To compute the standard error of $\hat{\beta}_1 + 21\hat{\beta}_2$, consider the F-statistic of

$$H_0\colon \beta_1 + 21\beta_2 = 0, \quad H_1\colon \beta_1 + 21\beta_2 \neq 0$$

  Since $F = 299.94$ and $\Delta\hat{Y} = 2.96$,

$$\mathsf{SE}(\Delta\hat{Y}) = \frac{|\Delta\hat{Y}|}{\sqrt{F}} = \frac{2.96}{\sqrt{299.94}} \approx 0.17$$

- A 95% confidence interval for the change in the expected value of Y is:

$$2.96 \pm 1.96 \times 0.17 = (2.63, 3.29)$$

Polynomial Regression

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.
$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0: \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1: \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.

$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0: \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1: \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.

$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1 : \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.

$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1 : \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.

$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0: \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1: \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Polynomial Regression Model

Polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots + \beta_r X_i^r + u_i$$

$r = 2$: quadratic regression model.

$r = 3$: cubic regression model.

- **Test the null hypothesis that the population regression function is linear:**

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_r = 0$$
$$H_1 : \text{not } H_0$$

- **Strategy to determine the degree of the polynomial regression:**

1. Pick a large value of r and estimate the polynomial regression.
2. Use the t-statistic to test the hypothesis $\beta_r = 0$.
3. If you cannot reject $\beta_r = 0$, estimate a polynomial regression of degree $r - 1$.
4. Continue until the coefficient on the highest power is statistically significant.

Cubic Effects On Test Scores

Test Scores and District Income

Cubic regression model:

- Estimate of cubic regression model relating test scores and income:

$$\begin{aligned}\widehat{TestScore} = \underset{(5.1)}{600.1} + &\underset{(0.71)}{5.02}\ Income - \underset{(0.029)}{0.096}\ Income^2 \\ + &\underset{(0.00035)}{0.00069}\ Income^3, \quad \bar{R}^2 = 0.555\end{aligned}$$

- Test $H_0: \beta_3 = 0, \quad H_1: \beta_3 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_3 - 0}{\text{SE}(\hat{\beta}_3)} = \frac{0.00069}{0.00035} \approx 1.97$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis at the 5% level — barely.

Test Scores and District Income

Cubic regression model:

- Estimate of cubic regression model relating test scores and income:

$$\begin{aligned}\widehat{TestScore} = \underset{(5.1)}{600.1} + &\underset{(0.71)}{5.02}\ Income - \underset{(0.029)}{0.096}\ Income^2 \\ + &\underset{(0.00035)}{0.00069}\ Income^3, \quad \bar{R}^2 = 0.555\end{aligned}$$

- Test $H_0: \beta_3 = 0, \quad H_1: \beta_3 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_3 - 0}{\text{SE}(\hat{\beta}_3)} = \frac{0.00069}{0.00035} \approx 1.97$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis at the 5% level — barely.

Test Scores and District Income

Cubic regression model:

- Estimate of cubic regression model relating test scores and income:

$$\begin{aligned}\widehat{TestScore} = \underset{(5.1)}{600.1} + \underset{(0.71)}{5.02}\ Income - \underset{(0.029)}{0.096}\ Income^2 \\ + \underset{(0.00035)}{0.00069}\ Income^3, \quad \bar{R}^2 = 0.555\end{aligned}$$

- Test $H_0: \beta_3 = 0, \quad H_1: \beta_3 \neq 0$

$$t^{\text{act}} = \frac{\hat{\beta}_3 - 0}{\text{SE}(\hat{\beta}_3)} = \frac{0.00069}{0.00035} \approx 1.97$$

- Since $t^{\text{act}} > 1.96$, we can reject the null hypothesis at the 5% level — barely.

Logarithmic Regression

Logarithm Function

- **Logarithm function:** $\ln(x)$ is the inverse of the exponential function e^x.
- **Exponential function:** e^x, for $e \approx 2.71828\ldots$
- **Properties:**
  The logarithmic function is defined on $x > 0$. It is strictly increasing with slope $1/x$: It is steeper for smaller values of x. Its limits are:

$$\ln(x) \to -\infty \text{ as } x \to 0$$
$$\ln(x) \to +\infty \text{ as } x \to +\infty$$

- **Other Useful properties:**

$$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x),$$
$$\ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x).$$

- The slope of the log function gives the growth rate: For a small change Δx,

$$\boxed{\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}}$$

Logarithm Function

- **Logarithm function:** $\ln(x)$ is the inverse of the exponential function e^x.
- **Exponential function:** e^x, for $e \approx 2.71828\ldots$
- **Properties:**
  The logarithmic function is defined on $x > 0$. It is strictly increasing with slope $1/x$: It is steeper for smaller values of x. Its limits are:

$$\ln(x) \to -\infty \text{ as } x \to 0$$
$$\ln(x) \to +\infty \text{ as } x \to +\infty$$

- **Other Useful properties:**

$$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x),$$
$$\ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x).$$

- The slope of the log function gives the growth rate: For a small change $\Delta\, x$,

$$\boxed{\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta\, x}{x}}$$

Logarithm Function

- **Logarithm function:** $\ln(x)$ is the inverse of the exponential function e^x.
- **Exponential function:** e^x, for $e \approx 2.71828\ldots$
- **Properties:**

  The logarithmic function is defined on $x > 0$. It is strictly increasing with slope $1/x$: It is steeper for smaller values of x. Its limits are:

  $$\ln(x) \to -\infty \text{ as } x \to 0$$
  $$\ln(x) \to +\infty \text{ as } x \to +\infty$$

- **Other Useful properties:**

  $$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x),$$
  $$\ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x).$$

- The slope of the log function gives the growth rate: For a small change Δx,

  $$\boxed{\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}}$$

Logarithm Function

- **Logarithm function:** $\ln(x)$ is the inverse of the exponential function e^x.
- **Exponential function:** e^x, for $e \approx 2.71828\ldots$
- **Properties:**

  The logarithmic function is defined on $x > 0$. It is strictly increasing with slope $1/x$: It is steeper for smaller values of x. Its limits are:

$$\ln(x) \to -\infty \text{ as } x \to 0$$
$$\ln(x) \to +\infty \text{ as } x \to +\infty$$

- **Other Useful properties:**

$$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x),$$
$$\ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x).$$

- The slope of the log function gives the growth rate: For a small change Δx,

$$\boxed{\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}}$$

Logarithm Function

- **Logarithm function:** $\ln(x)$ is the inverse of the exponential function e^x.
- **Exponential function:** e^x, for $e \approx 2.71828\ldots$
- **Properties:**
  The logarithmic function is defined on $x > 0$. It is strictly increasing with slope $1/x$: It is steeper for smaller values of x. Its limits are:

$$\ln(x) \to -\infty \text{ as } x \to 0$$
$$\ln(x) \to +\infty \text{ as } x \to +\infty$$

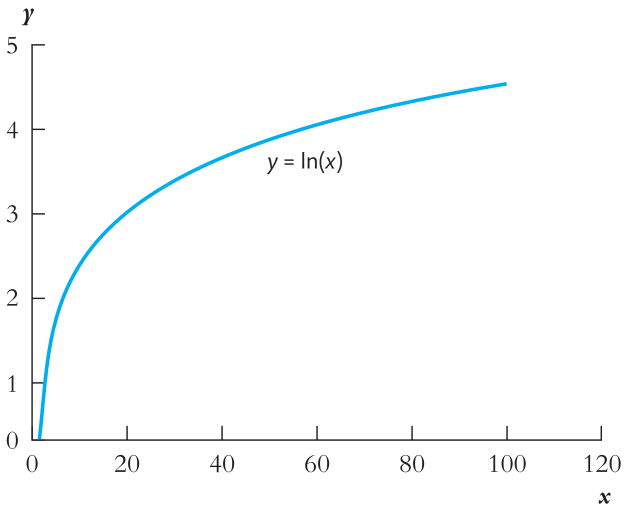
- **Other Useful properties:**

$$\ln(1/x) = -\ln(x), \quad \ln(ax) = \ln(a) + \ln(x),$$
$$\ln(x/a) = \ln(x) - \ln(a), \quad \ln(x^a) = a\ln(x).$$

- The slope of the log function gives the growth rate: For a small change $\Delta\, x$,

$$\boxed{\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta\, x}{x}}$$

Nonlinear Functions: The Logarithm



Log Regression Models

- **Linear-Log Model:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $0.01\beta_1$ change in Y.

- **Log-Linear Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

A one-unit change in X is associated with a $100\beta_1\%$ change in Y.

- **Log-Log Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $\beta_1\%$ change in Y.

- In the log-log model, β_1 is the elasticity of Y with respect to X.

Log Regression Models

- **Linear-Log Model:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $0.01\beta_1$ change in Y.

- **Log-Linear Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

A one-unit change in X is associated with a $100\beta_1\%$ change in Y.

- **Log-Log Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $\beta_1\%$ change in Y.

- In the log-log model, β_1 is the elasticity of Y with respect to X.

Log Regression Models

- **Linear-Log Model:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $0.01\beta_1$ change in Y.

- **Log-Linear Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

A one-unit change in X is associated with a $100\beta_1\%$ change in Y.

- **Log-Log Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

A one-percent change in X is associated with a $\beta_1\%$ change in Y.

- In the log-log model, β_1 is the elasticity of Y with respect to X.

Log Regression Models

- **Linear-Log Model:**

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

  A one-percent change in X is associated with a $0.01\beta_1$ change in Y.

- **Log-Linear Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

  A one-unit change in X is associated with a $100\beta_1\%$ change in Y.

- **Log-Log Model:**

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

  A one-percent change in X is associated with a $\beta_1\%$ change in Y.

- In the log-log model, β_1 is the elasticity of Y with respect to X.

Log Effects On Test Scores

Test Scores and District Income

Linear-Log regression model:

- Estimate of linear-log regression model relating test scores and income:

$$\underset{}{\widehat{TestScore}} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42}\ \ln(Income), \quad \bar{R}^2 = 0.561$$

- A 1% increase in income causes an increase in test scores of $0.01 \times 36.42 = 0.36$ points.
- Predicted difference in test scores for districts with average incomes of $\$10,000$ vs $\$11,000$:

$$\begin{aligned}\Delta\hat{Y} &= [557.8 + 36.42\ln(11)] - [557.8 + 36.42\ln(10)]\\ &= 36.42[\ln(11) - \ln(10)] \approx 3.47\end{aligned}$$

- Predicted difference in test scores for districts with average incomes of $\$40,000$ vs $\$41,000$:

$$\Delta\hat{Y} = 36.42[\ln(41) - \ln(40)] \approx 0.90$$

- A $\$1000$ increase in income has a larger effect on test scores in poor districts than it does in affluent districts.

Test Scores and District Income

Linear-Log regression model:

- Estimate of linear-log regression model relating test scores and income:

$$\widehat{TestScore} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42} \ \ln(Income), \quad \bar{R}^2 = 0.561$$

- A 1% increase in income causes an increase in test scores of $0.01 \times 36.42 = 0.36$ points.
- Predicted difference in test scores for districts with average incomes of $\$10,000$ vs $\$11,000$:

$$\begin{aligned}\Delta\hat{Y} &= [557.8 + 36.42\ln(11)] - [557.8 + 36.42\ln(10)] \\ &= 36.42[\ln(11) - \ln(10)] \approx 3.47\end{aligned}$$

- Predicted difference in test scores for districts with average incomes of $\$40,000$ vs $\$41,000$:

$$\Delta\hat{Y} = 36.42[\ln(41) - \ln(40)] \approx 0.90$$

- A $\$1000$ increase in income has a larger effect on test scores in poor districts than it does in affluent districts.

Test Scores and District Income

Linear-Log regression model:

- Estimate of linear-log regression model relating test scores and income:

$$\widehat{TestScore} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42} \ \ln(Income), \quad \bar{R}^2 = 0.561$$

- A 1% increase in income causes an increase in test scores of $0.01 \times 36.42 = 0.36$ points.
- Predicted difference in test scores for districts with average incomes of $\$10,000$ vs $\$11,000$:

$$\begin{aligned}\Delta\hat{Y} &= [557.8 + 36.42\ln(11)] - [557.8 + 36.42\ln(10)] \\ &= 36.42[\ln(11) - \ln(10)] \approx 3.47\end{aligned}$$

- Predicted difference in test scores for districts with average incomes of $\$40,000$ vs $\$41,000$:

$$\Delta\hat{Y} = 36.42[\ln(41) - \ln(40)] \approx 0.90$$

- A $\$1000$ increase in income has a larger effect on test scores in poor districts than it does in affluent districts.

Test Scores and District Income

Linear-Log regression model:

- Estimate of linear-log regression model relating test scores and income:

$$\widehat{TestScore} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42}\ \ln(Income), \quad \bar{R}^2 = 0.561$$

- A 1% increase in income causes an increase in test scores of $0.01 \times 36.42 = 0.36$ points.
- Predicted difference in test scores for districts with average incomes of $\$10,000$ vs $\$11,000$:

$$\begin{aligned}\Delta\hat{Y} &= [557.8 + 36.42\ln(11)] - [557.8 + 36.42\ln(10)] \\ &= 36.42[\ln(11) - \ln(10)] \approx 3.47\end{aligned}$$

- Predicted difference in test scores for districts with average incomes of $\$40,000$ vs $\$41,000$:

$$\Delta\hat{Y} = 36.42[\ln(41) - \ln(40)] \approx 0.90$$

- A $\$1000$ increase in income has a larger effect on test scores in poor districts than it does in affluent districts.

Test Scores and District Income

Linear-Log regression model:

- Estimate of linear-log regression model relating test scores and income:

$$\underset{}{\widehat{TestScore}} = \underset{(3.8)}{557.8} + \underset{(1.40)}{36.42}\ \ln(Income), \quad \bar{R}^2 = 0.561$$

- A 1% increase in income causes an increase in test scores of $0.01 \times 36.42 = 0.36$ points.
- Predicted difference in test scores for districts with average incomes of $\$10,000$ vs $\$11,000$:

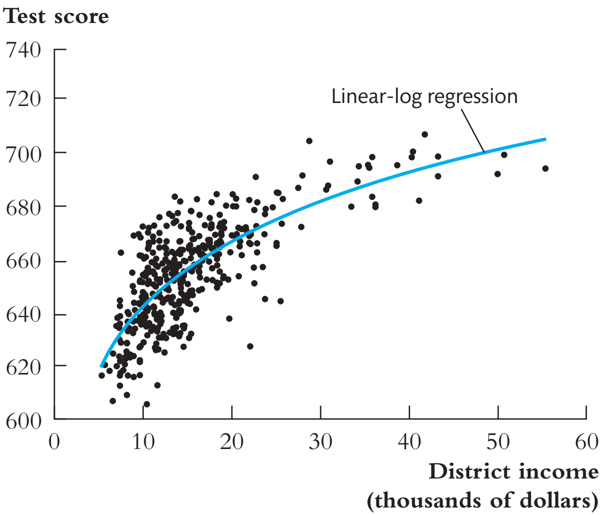
$$\begin{aligned}\Delta\hat{Y} &= [557.8 + 36.42\ln(11)] - [557.8 + 36.42\ln(10)] \\ &= 36.42[\ln(11) - \ln(10)] \approx 3.47\end{aligned}$$

- Predicted difference in test scores for districts with average incomes of $\$40,000$ vs $\$41,000$:

$$\Delta\hat{Y} = 36.42[\ln(41) - \ln(40)] \approx 0.90$$

- A $\$1000$ increase in income has a larger effect on test scores in poor districts than it does in affluent districts.

Linear-Log Regression



Test Scores and District Income

Log-Linear vs. Log-Log regression model:

- Estimate of log-linear regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284}\ Income, \quad \bar{R}^2 = 0.497$$

- A 1 dollar increase in income is associated with an increase in test scores of $100 \times 0.00284 = 0.284$ percent.
- Estimate of log-log regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554}\ \ln(Income), \quad \bar{R}^2 = 0.557$$

- A 1% increase in income is associated with an increase in test scores of 0.0554 percent.
- The log-log model appears to be a better fit than the log-linear model. But unfortunately the $\bar{R}^2$ cannot be used to compare the models, because they use different dependent variables.

Test Scores and District Income

Log-Linear vs. Log-Log regression model:

- Estimate of log-linear regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284} \; Income, \quad \bar{R}^2 = 0.497$$

- A 1 dollar increase in income is associated with an increase in test scores of $100 \times 0.00284 = 0.284$ percent.
- Estimate of log-log regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554} \; \ln(Income), \quad \bar{R}^2 = 0.557$$

- A 1% increase in income is associated with an increase in test scores of 0.0554 percent.
- The log-log model appears to be a better fit than the log-linear model. But unfortunately the $\bar{R}^2$ cannot be used to compare the models, because they use different dependent variables.

Test Scores and District Income

Log-Linear vs. Log-Log regression model:

- Estimate of log-linear regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284}\ Income, \quad \bar{R}^2 = 0.497$$

- A 1 dollar increase in income is associated with an increase in test scores of $100 \times 0.00284 = 0.284$ percent.
- Estimate of log-log regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554}\ \ln(Income), \quad \bar{R}^2 = 0.557$$

- A 1% increase in income is associated with an increase in test scores of 0.0554 percent.
- The log-log model appears to be a better fit than the log-linear model. But unfortunately the $\bar{R}^2$ cannot be used to compare the models, because they use different dependent variables.

Test Scores and District Income

Log-Linear vs. Log-Log regression model:

- Estimate of log-linear regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284}\ Income, \quad \bar{R}^2 = 0.497$$

- A 1 dollar increase in income is associated with an increase in test scores of $100 \times 0.00284 = 0.284$ percent.
- Estimate of log-log regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554}\ \ln(Income), \quad \bar{R}^2 = 0.557$$

- A 1% increase in income is associated with an increase in test scores of 0.0554 percent.
- The log-log model appears to be a better fit than the log-linear model. But unfortunately the $\bar{R}^2$ cannot be used to compare the models, because they use different dependent variables.

Test Scores and District Income

Log-Linear vs. Log-Log regression model:

- Estimate of log-linear regression model relating test scores and income:

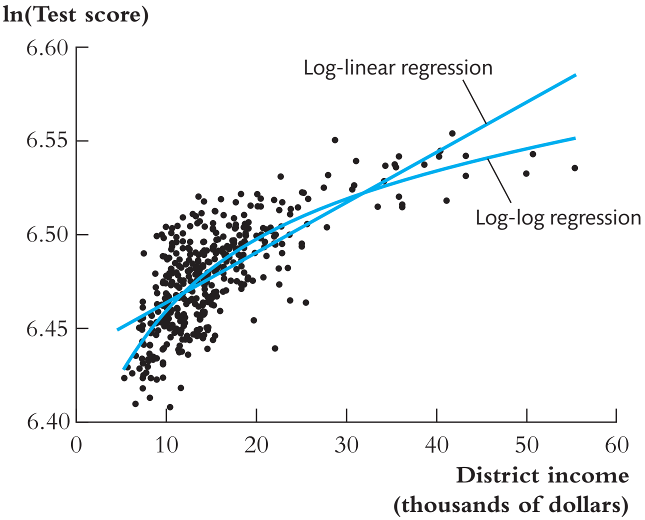
$$\widehat{\ln(TestScore)} = \underset{(0.003)}{6.439} + \underset{(0.00018)}{0.00284} \; Income, \quad \bar{R}^2 = 0.497$$

- A 1 dollar increase in income is associated with an increase in test scores of $100 \times 0.00284 = 0.284$ percent.
- Estimate of log-log regression model relating test scores and income:

$$\widehat{\ln(TestScore)} = \underset{(0.006)}{6.336} + \underset{(0.0021)}{0.0554} \; \ln(Income), \quad \bar{R}^2 = 0.557$$

- A 1% increase in income is associated with an increase in test scores of 0.0554 percent.
- The log-log model appears to be a better fit than the log-linear model. But unfortunately the $\bar{R}^2$ cannot be used to compare the models, because they use different dependent variables.

Log-Linear and Log-Log Regressions



Log-Cubic Effects On Test Scores

Test Scores and District Income

Linear-Log-Polynomial regression model:

- Estimate of linear-log-polynomial regression model relating test scores and income:

$$\begin{aligned}\widehat{TestScore} = \underset{(79.4)}{486.1} + \underset{(87.9)}{113.4}\ \ln(Income) - \underset{(31.7)}{26.9}\ [\ln(Income)]^2 \\ - \underset{(3.74)}{3.06}\ [\ln(Income)]^3, \quad \bar{R}^2 = 0.560\end{aligned}$$

- The null hypothesis that the true coefficient on the cubic term is zero cannot be rejected at the 10% significance level.
- The F-statistic for the joint hypothesis that the true coefficients on the quadratic and cubic terms are both zero cannot be rejected at the 10% level.
- The cubic logarithmic model is not, statistically speaking, an improvement over the linear-log model.

Test Scores and District Income

Linear-Log-Polynomial regression model:

- Estimate of linear-log-polynomial regression model relating test scores and income:

$$\begin{aligned}\widehat{TestScore} = \underset{(79.4)}{486.1} + \underset{(87.9)}{113.4}\ \ln(Income) - \underset{(31.7)}{26.9}\ [\ln(Income)]^2 \\ - \underset{(3.74)}{3.06}\ [\ln(Income)]^3, \quad \bar{R}^2 = 0.560\end{aligned}$$

- The null hypothesis that the true coefficient on the cubic term is zero cannot be rejected at the 10% significance level.
- The F-statistic for the joint hypothesis that the true coefficients on the quadratic and cubic terms are both zero cannot be rejected at the 10% level.
- The cubic logarithmic model is not, statistically speaking, an improvement over the linear-log model.

Test Scores and District Income

Linear-Log-Polynomial regression model:

- Estimate of linear-log-polynomial regression model relating test scores and income:

$$\begin{aligned}
\widehat{TestScore} = \underset{(79.4)}{486.1} + \underset{(87.9)}{113.4}\ \ln(Income) - \underset{(31.7)}{26.9}\ [\ln(Income)]^2 \\
- \underset{(3.74)}{3.06}\ [\ln(Income)]^3, \quad \bar{R}^2 = 0.560
\end{aligned}$$

- The null hypothesis that the true coefficient on the cubic term is zero cannot be rejected at the 10% significance level.
- The F-statistic for the joint hypothesis that the true coefficients on the quadratic and cubic terms are both zero cannot be rejected at the 10% level.
- The cubic logarithmic model is not, statistically speaking, an improvement over the linear-log model.

Test Scores and District Income

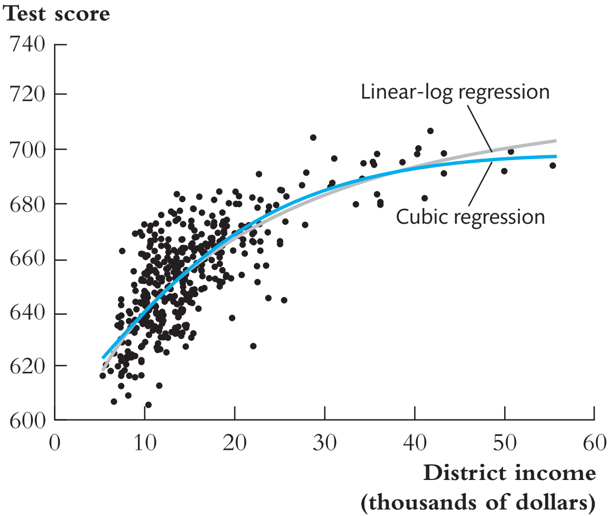
Linear-Log-Polynomial regression model:

- Estimate of linear-log-polynomial regression model relating test scores and income:

$$
\begin{aligned}
\widehat{TestScore} = \underset{(79.4)}{486.1} + \underset{(87.9)}{113.4} \ln(Income) - \underset{(31.7)}{26.9} [\ln(Income)]^2 \\
- \underset{(3.74)}{3.06} [\ln(Income)]^3, \quad \bar{R}^2 = 0.560
\end{aligned}
$$

- The null hypothesis that the true coefficient on the cubic term is zero cannot be rejected at the 10% significance level.
- The F-statistic for the joint hypothesis that the true coefficients on the quadratic and cubic terms are both zero cannot be rejected at the 10% level.
- The cubic logarithmic model is not, statistically speaking, an improvement over the linear-log model.

Linear-Log and Cubic Regressions



Interaction Terms

Interactions Between Independent Variables

- **Interaction between two binary variables:**

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 \, D_{1i} \cdot D_{2i} + u_i$$

where $D_{1i} \cdot D_{2i}$ is called an interaction term.

- **Interpretation:**
  1. Compute the expected values of Y for each possible case described by the binary variables
  2. Compare the expected values in each case.
  3. Each coefficient can be expressed either as an expected value or as the difference between two or more expected values.

Interactions Between Independent Variables

- **Interaction between two binary variables:**

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 \, D_{1i} \cdot D_{2i} + u_i$$

  where $D_{1i} \cdot D_{2i}$ is called an interaction term.

- **Interpretation:**
  1. Compute the expected values of Y for each possible case described by the binary variables.
  2. Compare the expected values in each case.
  3. Each coefficient can be expressed either as an expected value or as the difference between two or more expected values.

Interactions Between Independent Variables

- **Interaction between two binary variables:**

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3\, D_{1i} \cdot D_{2i} + u_i$$

  where $D_{1i} \cdot D_{2i}$ is called an interaction term.

- **Interpretation:**
  1. Compute the expected values of Y for each possible case described by the binary variables.
  2. Compare the expected values in each case.
  3. Each coefficient can be expressed either as an expected value or as the difference between two or more expected values.

Interactions Between Independent Variables

- **Interaction between two binary variables:**

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3\, D_{1i} \cdot D_{2i} + u_i$$

  where $D_{1i} \cdot D_{2i}$ is called an interaction term.

- **Interpretation:**
  1. Compute the expected values of Y for each possible case described by the binary variables.
  2. Compare the expected values in each case.
  3. Each coefficient can be expressed either as an expected value or as the difference between two or more expected values.

Interactions Between Independent Variables

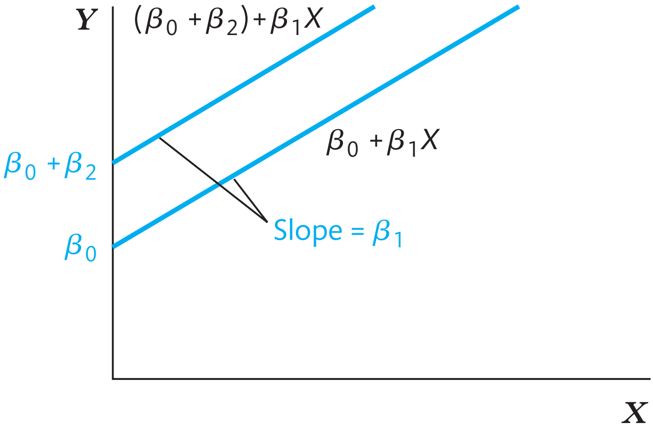
- **Interaction between two binary variables:**

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 \, D_{1i} \cdot D_{2i} + u_i$$

  where $D_{1i} \cdot D_{2i}$ is called an interaction term.

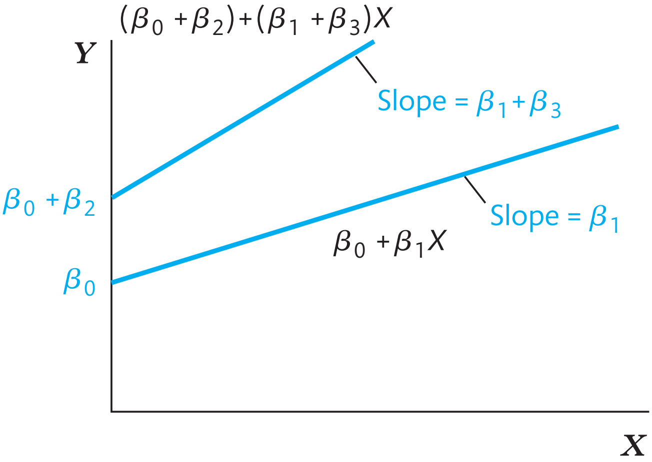
- **Interpretation:**
  1. Compute the expected values of Y for each possible case described by the binary variables.
  2. Compare the expected values in each case.
  3. Each coefficient can be expressed either as an expected value or as the difference between two or more expected values.

Regressions with Binary and Continuous Variables



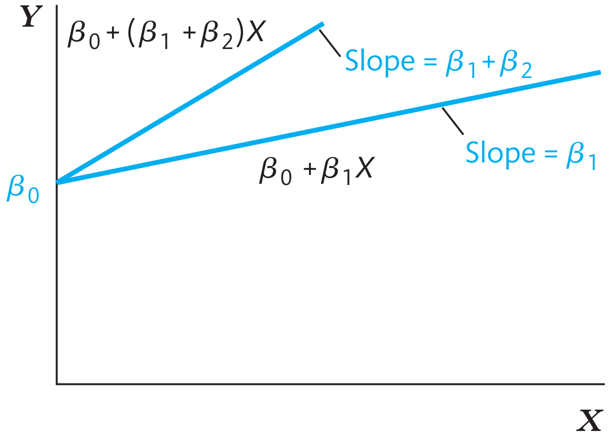
(a) Different intercepts, same slope

Regressions with Binary and Continuous Variables



(b) Different intercepts, different slopes

Regressions with Binary and Continuous Variables



(c) Same intercept, different slopes

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3\, X_i \cdot D_i + u_i$$

- **Interpretation:**

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 \, X_i \cdot D_i + u_i$$

- **Interpretation:**

  1. If $D_i = 0$, the population regression function is $\beta_0 + \beta_1 X_i$.
  2. If $D_i = 1$, the population regression function is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i$.
  3. The difference between the two intercepts is β_2.
  4. The difference between the two slopes is β_3.

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 \, X_i \cdot D_i + u_i$$

- **Interpretation:**
  1. If $D_i = 0$, the population regression function is $\beta_0 + \beta_1 X_i$.
  2. If $D_i = 1$, the population regression function is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i$.
  3. The difference between the two intercepts is β_2.
  4. The difference between the two slopes is β_3.

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3\, X_i \cdot D_i + u_i$$

- **Interpretation:**
  1. If $D_i = 0$, the population regression function is $\beta_0 + \beta_1 X_i$.
  2. If $D_i = 1$, the population regression function is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i$.
  3. The difference between the two intercepts is β_2.
  4. The difference between the two slopes is β_3.

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3\, X_i \cdot D_i + u_i$$

- **Interpretation:**
  1. If $D_i = 0$, the population regression function is $\beta_0 + \beta_1 X_i$.
  2. If $D_i = 1$, the population regression function is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i$.
  3. The difference between the two intercepts is β_2.
  4. The difference between the two slopes is β_3.

Interactions Between Independent Variables

- **Interaction between a continuous and a binary variable:**

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 \, X_i \cdot D_i + u_i$$

- **Interpretation:**
  1. If $D_i = 0$, the population regression function is $\beta_0 + \beta_1 X_i$.
  2. If $D_i = 1$, the population regression function is $(\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i$.
  3. The difference between the two intercepts is β_2.
  4. The difference between the two slopes is β_3.

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3\, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.

- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.

- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3\, X_2$$

and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3\, X_2)\Delta X_1 + (\beta_2 + \beta_3\, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3\, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.
- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.
- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3\, X_2$$

  and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3\, X_2)\Delta X_1 + (\beta_2 + \beta_3\, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.

- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.

- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 \, X_2$$

  and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3 \, X_2)\Delta X_1 + (\beta_2 + \beta_3 \, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3\, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.
- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.
- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3\, X_2$$

  and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3\, X_2)\Delta X_1 + (\beta_2 + \beta_3\, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.
- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.
- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 \, X_2$$

  and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3 \, X_2)\Delta X_1 + (\beta_2 + \beta_3 \, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Interactions Between Independent Variables

- **Interaction between two continuous variables:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3\, X_{1i} \cdot X_{2i} + u_i$$

- **Interpretation:** The interaction term
  1. allows the effect of a change in X_1 to depend on X_2.
  2. allows the effect of a change in X_2 to depend on the value of X_1.
- The coefficient on $X_1 \cdot X_2$ is the effect of a one-unit increase in X_1 and X_2, beyond the sum of the individual effects of a unit increase in X_1 alone and a unit increase in X_2 alone.
- The effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3\, X_2$$

  and likewise for ΔX_2. If X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y is:

$$\Delta Y = (\beta_1 + \beta_3\, X_2)\Delta X_1 + (\beta_2 + \beta_3\, X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Return to Education and the Gender Gap

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings – not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - For 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings — not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - For 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings — not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings — not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings – not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - for 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings – not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - for 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings – not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - for 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap

The Return to Education revisited:

- A multiple regression analysis that controls for determinants of earnings that, if omitted, could cause omitted variable bias, and that uses a nonlinear functional form.
- The next Table summarizes regressions estimated using data on $47,233$ full-time workers, ages 30 through 64, from the Current Population Survey (CPS).
- The dependent variable is the logarithm of hourly earnings, so an additional year of education is associated with a constant percentage increase in earnings – not a dollar increase.
- The estimated economic return to education in regression (4) is 11.14% for each year of education for men and $0.1114 + 0.0082 = 11.96\%$ for women.
- Because the regression functions for men and women have different slopes, the gender gap depends on the years of education.
  - For 12 years of education, the gender gap is 27.0% $(0.0082 \times 12 - 0.368)$.
  - for 16 years of education, the gender gap is 23.7% $(0.0082 \times 16 - 0.368)$.
- Labor economists estimate the return to education between 8% and 11%, depending on the quality of the education received.

Return to Education and Gender Gap: United States, 2015

Dependent variable: logarithm of *Hourly earnings*.

Regressor	(1)	(2)	(3)	(4)
Years of education	0.1056 (0.0009)	0.1089 (0.0009)	0.1063 (0.0018)	0.1114 (0.0013)
Female		−0.252 (0.005)	−0.342 (0.026)	−0.368 (0.026)
Female × *Years of education*			0.0063 (0.0018)	0.0082 (0.0018)
Potential experience				0.0147 (0.0013)
Potential experience^2				−0.000183 (0.000024)
a. Regional control variables?	No	No	No	Yes
95% confidence interval for return to education				
Combined men & women	[0.104, 0.107]	[0.107, 0.111]		
For men			[0.104, 0.109]	[0.109, 0.114]
For women			[0.110, 0.115]	[0.117, 0.122]
$\overline{R}^2$	0.209	0.251	0.251	0.262

Return to Education and Gender Gap

The Gender Gap revisited:

1. The omission of sex in regression (1) does not result in substantial omitted variable bias: Even though sex enters regression (2) significantly and with a large coefficient, sex and years of education are nearly uncorrelated: On average, men and women have nearly the same levels of education.
2. The returns to education are economically and statistically significantly different for men and women: In regression (3), the t-statistic testing the hypothesis that they are the same is 3.42. The confidence interval is tight: the return to education is precisely estimated both for men and for women.
3. Regression (4) controls for the region of the country in which the individual lives, to address potential omitted variable bias that might arise if years of education differ systematically by region. Controlling for region makes a small difference to the estimated coefficients on the education terms relative to those reported in regression (3).
4. Regression (4) controls for the potential experience of the worker, as measured by years since completion of schooling. The estimated coefficients imply a declining marginal value for each year of potential experience.

The Gender Gap revisited:

1. The omission of sex in regression (1) does not result in substantial omitted variable bias: Even though sex enters regression (2) significantly and with a large coefficient, sex and years of education are nearly uncorrelated: On average, men and women have nearly the same levels of education.
2. The returns to education are economically and statistically significantly different for men and women: In regression (3), the t-statistic testing the hypothesis that they are the same is 3.42. The confidence interval is tight: the return to education is precisely estimated both for men and for women.
3. Regression (4) controls for the region of the country in which the individual lives, to address potential omitted variable bias that might arise if years of education differ systematically by region. Controlling for region makes a small difference to the estimated coefficients on the education terms relative to those reported in regression (3).
4. Regression (4) controls for the potential experience of the worker, as measured by years since completion of schooling. The estimated coefficients imply a declining marginal value for each year of potential experience.

Return to Education and Gender Gap

The Gender Gap revisited:

1. The omission of sex in regression (1) does not result in substantial omitted variable bias: Even though sex enters regression (2) significantly and with a large coefficient, sex and years of education are nearly uncorrelated: On average, men and women have nearly the same levels of education.
2. The returns to education are economically and statistically significantly different for men and women: In regression (3), the t-statistic testing the hypothesis that they are the same is 3.42. The confidence interval is tight: the return to education is precisely estimated both for men and for women.
3. Regression (4) controls for the region of the country in which the individual lives, to address potential omitted variable bias that might arise if years of education differ systematically by region. Controlling for region makes a small difference to the estimated coefficients on the education terms relative to those reported in regression (3).
4. Regression (4) controls for the potential experience of the worker, as measured by years since completion of schooling. The estimated coefficients imply a declining marginal value for each year of potential experience.

Return to Education and Gender Gap

The Gender Gap revisited:

1. The omission of sex in regression (1) does not result in substantial omitted variable bias: Even though sex enters regression (2) significantly and with a large coefficient, sex and years of education are nearly uncorrelated: On average, men and women have nearly the same levels of education.
2. The returns to education are economically and statistically significantly different for men and women: In regression (3), the t-statistic testing the hypothesis that they are the same is 3.42. The confidence interval is tight: the return to education is precisely estimated both for men and for women.
3. Regression (4) controls for the region of the country in which the individual lives, to address potential omitted variable bias that might arise if years of education differ systematically by region. Controlling for region makes a small difference to the estimated coefficients on the education terms relative to those reported in regression (3).
4. Regression (4) controls for the potential experience of the worker, as measured by years since completion of schooling. The estimated coefficients imply a declining marginal value for each year of potential experience.

Summary

Summary

- In a nonlinear regression, the slope of the population regression function depends on the value of one or more of the independent variables.
- The effect on Y of a change in the independent variables can be computed by evaluating the regression function at two values of the independent variables.
- A polynomial regression includes powers of X as regressors. A quadratic regression includes X and X^2, and a cubic regression includes X, X^2, and X^3.
- Small changes in logarithms can be interpreted as proportional or percentage changes in a variable. Regressions involving logarithms are used to estimate proportional changes and elasticities.
- The product of two variables is called an interaction term. When interaction terms are included as regressors, they allow the regression slope of one variable to depend on the value of another variable.

Summary

- In a nonlinear regression, the slope of the population regression function depends on the value of one or more of the independent variables.
- The effect on Y of a change in the independent variables can be computed by evaluating the regression function at two values of the independent variables.
- A polynomial regression includes powers of X as regressors. A quadratic regression includes X and X^2, and a cubic regression includes X, X^2, and X^3.
- Small changes in logarithms can be interpreted as proportional or percentage changes in a variable. Regressions involving logarithms are used to estimate proportional changes and elasticities.
- The product of two variables is called an interaction term. When interaction terms are included as regressors, they allow the regression slope of one variable to depend on the value of another variable.

Summary

- In a nonlinear regression, the slope of the population regression function depends on the value of one or more of the independent variables.
- The effect on Y of a change in the independent variables can be computed by evaluating the regression function at two values of the independent variables.
- A polynomial regression includes powers of X as regressors. A quadratic regression includes X and X^2, and a cubic regression includes X, X^2, and X^3.
- Small changes in logarithms can be interpreted as proportional or percentage changes in a variable. Regressions involving logarithms are used to estimate proportional changes and elasticities.
- The product of two variables is called an interaction term. When interaction terms are included as regressors, they allow the regression slope of one variable to depend on the value of another variable.

Summary

- In a nonlinear regression, the slope of the population regression function depends on the value of one or more of the independent variables.
- The effect on Y of a change in the independent variables can be computed by evaluating the regression function at two values of the independent variables.
- A polynomial regression includes powers of X as regressors. A quadratic regression includes X and X^2, and a cubic regression includes X, X^2, and X^3.
- Small changes in logarithms can be interpreted as proportional or percentage changes in a variable. Regressions involving logarithms are used to estimate proportional changes and elasticities.
- The product of two variables is called an interaction term. When interaction terms are included as regressors, they allow the regression slope of one variable to depend on the value of another variable.

Summary

- In a nonlinear regression, the slope of the population regression function depends on the value of one or more of the independent variables.
- The effect on Y of a change in the independent variables can be computed by evaluating the regression function at two values of the independent variables.
- A polynomial regression includes powers of X as regressors. A quadratic regression includes X and X^2, and a cubic regression includes X, X^2, and X^3.
- Small changes in logarithms can be interpreted as proportional or percentage changes in a variable. Regressions involving logarithms are used to estimate proportional changes and elasticities.
- The product of two variables is called an interaction term. When interaction terms are included as regressors, they allow the regression slope of one variable to depend on the value of another variable.

Problems & Applications

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 8, Exercise 3.

After reading this chapter's analysis of test scores and class size, an educator comments, "In my experience, student performance depends on class size, but not in the way your regressions say. Rather, students do well when class size is less than 20 students and do very poorly when class size is greater than 25. There are no gains from reducing class size below 20 students, the relationship is constant in the intermediate region between 20 and 25 students, and there is no loss to increasing class size when it is already greater than 25." The educator is describing a threshold effect, in which performance is constant for class sizes less than 20, jumps and is constant for class sizes between 20 and 25, and then jumps again for class sizes greater than 25. To model these threshold effects, define the binary variables:

$$STRsmall = 1 \text{ if } STR < 20 \text{ and } STRsmall = 0 \text{ otherwise;}$$
$$STRmoderate = 1 \text{ if } 20 \leq STR \leq 25 \text{ and } STRmoderate = 0 \text{ otherwise;}$$
$$STRlarge = 1 \text{ if } STR > 25 \text{ and } STRlarge = 0 \text{ otherwise.}$$

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 8, Exercise 3.

1. Consider the regression $TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRlarge_i + u_i$. Sketch the regression function relating $TestScore$ to STR for hypothetical values of the regression coefficients that are consistent with the educator's statement.
2. A researcher tries to estimate the regression $TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRmoderate_i + \beta_3 STRlarge_i + u_i$ and finds that the software gives an error message. Why?

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 8, Exercise 3.

1. Consider the regression $TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRlarge_i + u_i$. Sketch the regression function relating $TestScore$ to STR for hypothetical values of the regression coefficients that are consistent with the educator's statement.
2. A researcher tries to estimate the regression $TestScore_i = \beta_0 + \beta_1 STRsmall_i + \beta_2 STRmoderate_i + \beta_3 STRlarge_i + u_i$ and finds that the software gives an error message. Why?