Topic 8

Nonlinear Regression

1) Modeling nonlinear regression

- In the linear regression model, we assume that the conditional expectation of Y_i given $X_{1i}, \ldots, X_{ki}$ is a linear function of the regressors. That is,

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$ where $E(u_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$

Equivalently, $E(Y_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$

- In nonlinear regression models, we allow this conditional expectation to be a nonlinear function of the regressors.

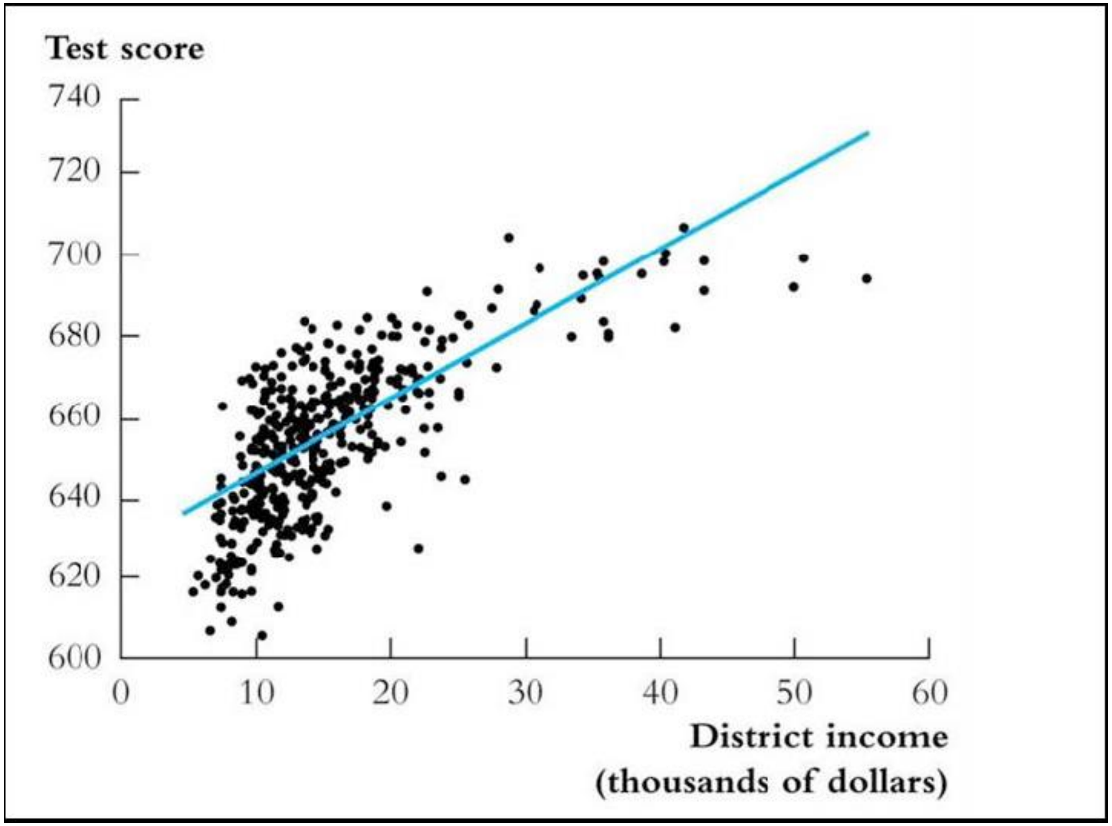
$Y_i = f(X_{1i}, X_{2i}, \ldots, X_{ki}) + u_i$ where $E(u_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$

Equivalently, $E(Y_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = f(X_{1i}, X_{2i}, \ldots, X_{ki})$

where *f is some nonlinear function.*

1) Modeling nonlinear regression

- The data in the scatterplot below shows test scores and district income in 420 Californian school districts in 1999.
- Does the relationship between test scores and district income appear linear?



1) Modeling nonlinear regression

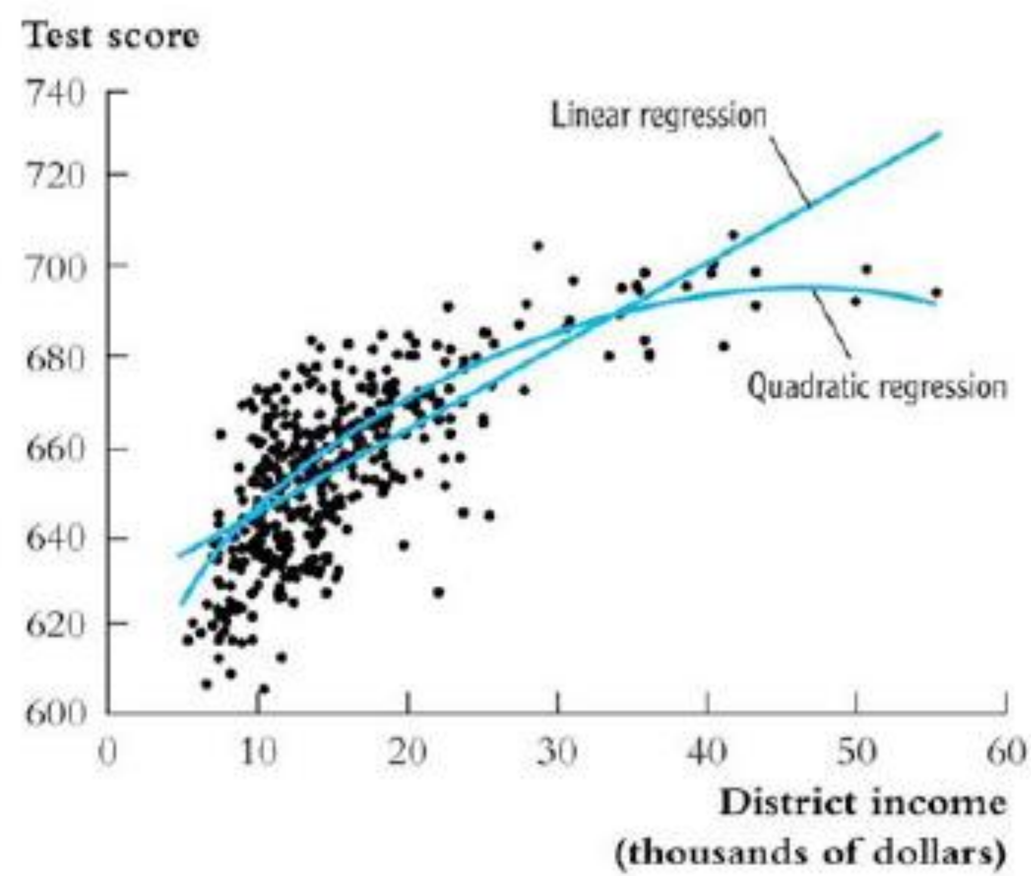
- If the relationship between test scores and district income is not linear, how should we model it?
- It appears that the effect upon test scores of increasing income gets smaller as income gets larger. We could try to model this effect using a quadratic regression.
- A quadratic regression model relating test scores (Y_i) to income (X_i) is

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

- This is simply a linear regression where we include X_{2i} as an additional regressor. We can estimate the parameters of the model by OLS in the usual way.

1) Modeling nonlinear regression

➤ The quadratic regression model appears to fit the data better than the linear regression model.



1) Modeling nonlinear regression

- Generate income square: avgincsquare=avginc^2

```
. gen avgincsquare=avginc^2

. sum testscr avginc avgincsquare
```

Variable	Obs	Mean	Std. Dev.	Min	Max
testscr	420	654.1565	19.05335	605.55	706.75
avginc	420	15.31659	7.22589	5.335	55.328
avgincsquare	420	286.687	351.2045	28.46222	3061.188

1) Modeling nonlinear regression


```
. reg testscr avginc avgincsquare, r

Linear regression                                      Number of obs =     420
                                                       F(  2,    417) =  428.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5562
                                                       Root MSE      =  12.724
```

testscr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
avginc	3.850995	.2680941	14.36	0.000	3.32401	4.377979
avgincsquare	-.0423085	.0047803	-8.85	0.000	-.051705	-.0329119
_cons	607.3017	2.901754	209.29	0.000	601.5978	613.0056

1) Modeling nonlinear regression

➢ We can formally test the null hypothesis of linearity against the alternative hypothesis of nonlinearity by examining the t-statistic on the estimated coefficient $\hat{\beta}_2$. In fact, we have

$$\begin{aligned} \hat{\beta}_2 &= -0.0423 \\ \hat{\sigma}_{\hat{\beta}_2} &= 0.0048, \end{aligned}$$

$$\text{so } t = \hat{\beta}_2 / \hat{\sigma}_{\hat{\beta}_2} = -8.85.$$

➢ The p-value corresponding to this t-statistic is less than 0.001, so we reject the null hypothesis $H_0 : \beta_2 = 0$.

➢ This means we reject the null hypothesis of linearity. The quadratic term is important.

1) Modeling nonlinear regression

- In the linear regression model, the interpretation of the regression coefficients $\beta_1, \ldots, \beta_k$ is simple. β_1 represents the expected change in Y if we increase X_1 by one unit while holding $X_2, \ldots, X_k$ constant.
- In our quadratic regression model, this interpretation makes no sense. β_1 cannot be the expected change in Y if we increase X while holding X^2 constant. This is nonsense because X^2 will always change when we increase X.
- In our quadratic regression, the expected change in Y if we increase X by one unit depends on the level of X. When X is large, the effect upon Y of changing X is smaller than when X is small.

1) Modeling nonlinear regression

- Lets go back to our general nonlinear regression model:

$$Y_i = f(X_{1i}, \ldots, X_{ki}) + u_i.$$

- In general, we can ask ourselves the question: What is the expected change in Y if X_1 increases by ΔX_1 while $X_2, \ldots, X_k$ stay the same?
- The answer is given by the equation

$$E(\Delta Y) = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k)$$

- If $\hat{f}$ is our estimated nonlinear regression function, then our prediction of the change in Y if X_1 increases by ΔX_1, while $X_2, \ldots, X_k$ stay the same, is
- $$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots, X_k) - \hat{f}(X_1, X_2, \ldots, X_k).$$

1) Modeling nonlinear regression

- For example, suppose we want to use our quadratic regression model to estimate the effect upon test scores of an increase in district income from $10,000 to $11,000.
- We have

$$
\begin{aligned}
&\Delta \hat{Y} \\
&= (\hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2) \\
&= \hat{\beta}_1 + (11^2 - 10^2)\hat{\beta}_2 = 3.85 + 21 \times (-0.0423) = 2.96.
\end{aligned}
$$

- What about the effect upon test scores of an increase in district income from $40,000 to $41,000?

$$
\begin{aligned}
&\Delta \hat{Y} \\
&= (\hat{\beta}_0 + \hat{\beta}_1 \times 41 + \hat{\beta}_2 \times 41^2) - (\hat{\beta}_0 + \hat{\beta}_1 \times 40 + \hat{\beta}_2 \times 40^2) \\
&= \hat{\beta}_1 + (41^2 - 40^2)\hat{\beta}_2 = 3.85 + 81 \times (-0.0423) = 0.42.
\end{aligned}
$$

2) Nonlinear functions of a single independent variable

- Our nonlinear regression model for the test score data involved a quadratic function of X. What about higher order polynomials?
- The *polynomial regression model* of degree r is given by

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i.$$

- When $r = 2$, we have the quadratic regression model discussed earlier. When $r = 3$ we have a cubic regression model, and when $r = 4$ we have a quartic regression model.
- We can estimate the parameters of the polynomial regression model simply by regressing Y_i on the powers of X_i using OLS.

2) Nonlinear functions of a single independent variable

- It can be difficult to choose the degree r to use with any given data set.
- Here is one possible way to choose *r* :

1. Start by choosing a maximum value of *r* - say 2,3 or 4 – and estimate the polynomial regression model for that *r.*

2. Use the *t*-statistic on the coefficient $\hat{\beta}_r$ to test the null hypothesis $H_0 : \beta_r = 0$. If you reject the null hypothesis, use the polynomial regression model of degree *r* for your analysis. Otherwise, go to step 3.

3. Subtract one from your value of *r* . Re-estimate the polynomial regression model by OLS, and return to step 2.

- This approach to model selection is known as the *general-to-specific* approach.

- For our test score and income data, this algorithm (with *r* initialized at 6) leads to a fifth order polynomial regression model if we use tests with a 5% size, and a fourth order polynomial regression model if we use tests with a 1% size.

2) Nonlinear functions of a single independent variable

- There are some aspects of the polynomial regression model that one should be wary of.
- In particular, using a polynomial regression function to extrapolate beyond the range of the data can be very problematic.
- In our test score example, the quadratic regression model implies that, when district income becomes very large, test scores start to decline. This may not seem plausible.

2) Nonlinear functions of a single independent variable

- When does the effect upon test scores of increasing income in the quadratic regression model start to become negative?
- Differentiating our quadratic regression function, we obtain

$$\frac{df(x)}{dx} = \frac{d}{dx}(\beta_0 + \beta_1 x + \beta_2 x^2) = \beta_1 + 2\beta_2 x.$$

- Setting the derivative equal to zero gives $x = -\beta_1 / 2\beta_2$. Since $\hat{\beta}_1 = 3.85$ nd $\hat{\beta}_2 = -0.0423$, we find that the marginal effect of increasing income is zero when $X = 45.5$.
- This means that, in our quadratic regression model, increasing income beyond $45,500 is predicted to have a negative effect upon test scores. This negative effect gets more extreme as income continues to rise.

2) Nonlinear functions of a single independent variable

- Another way to model a nonlinear relationship between the variables Y and X is to take the natural logarithm of one or both variables.
- The logarithm function is an increasing function. That is, $\ln b > \ln a$ when $b > a$. This property can be useful in applications like our test score example.
- In general, we use the logarithm function when it makes more sense to think about *percentage* changes in a variable, rather than numerical changes.
- For instance, the price elasticity of demand for some product is defined as the percentage change in demand that would occur as a result of a 1% change in price.
- Changes in many economic variables are typically described in percentage terms - price levels, GDP, unemployment, etc.

2) Nonlinear functions of a single independent variable

- The relationship between the logarithm function and percentage changes is due to the following approximation:

$$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x} \text{ when } \frac{\Delta x}{x} \text{ is small.}$$

- So, if we increase *x* by 1%, we have $(\Delta x)/x = 0.01$, and so $\ln(x + \Delta x) \approx \ln(x) + 0.01$. That is, a 1% increase in *x* causes ln(*x*) to increase by approximately 0.01.

2) Nonlinear functions of a single independent variable

- The *linear-log* regression model is

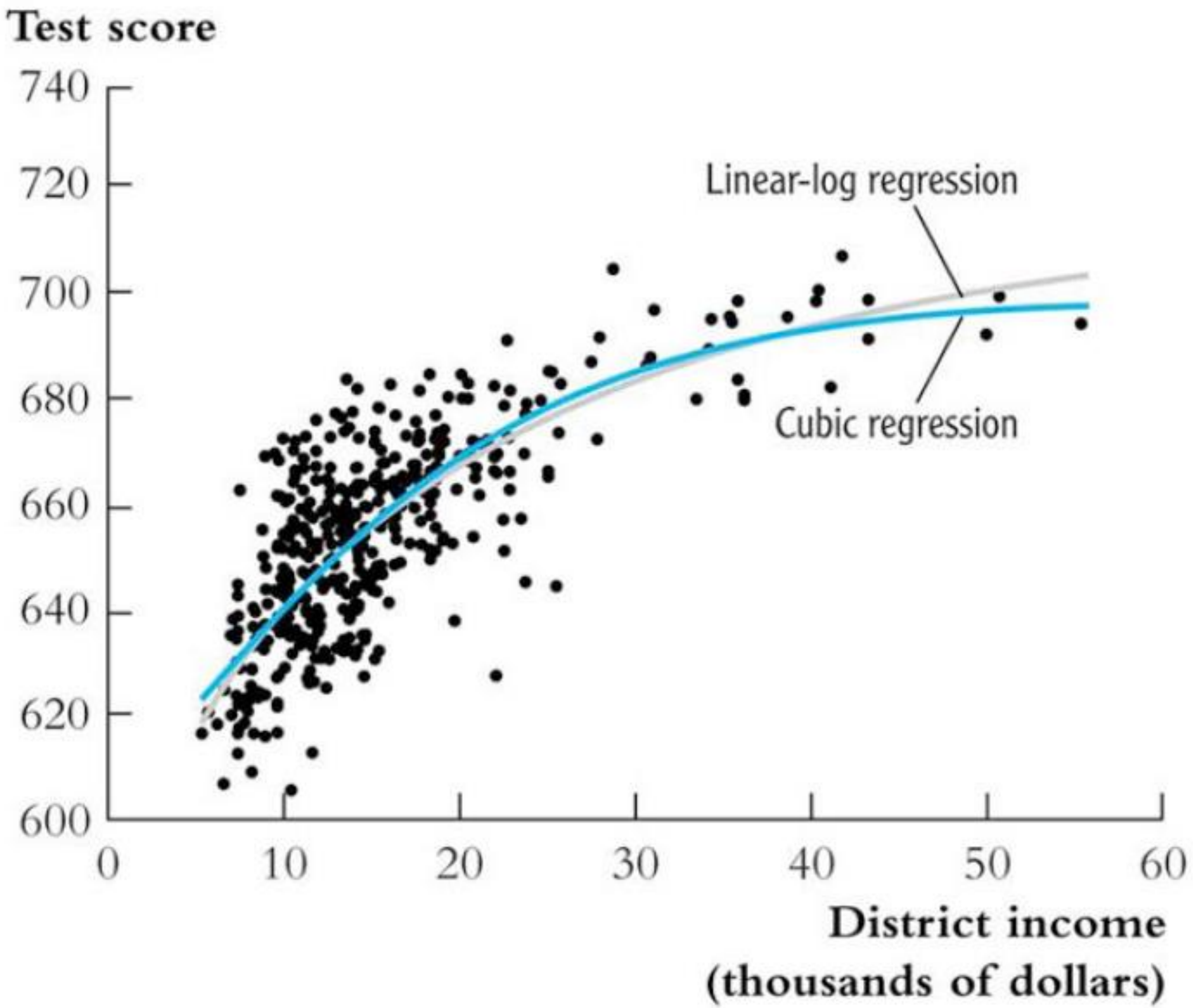
$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i.$$

- In this model, the expected change in Y_i if we increase X_i by 1% is approximately $0.01 \times \beta_1$

- We can estimate the parameters β_0 and β_1 by regressing Y_i on $\ln(X_i)$ using OLS.

2) Nonlinear functions of a single independent variable

➢



2) Nonlinear functions of a single independent variable

- The *log-linear* regression model is

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

- Note that we can rewrite this equation as

$$Y_i = Ae^{\beta_1 X_i} e^{u_i},$$

where $A = e^{\beta_0}$.

- We can estimate β_0 and β_1 by regressing ln(Y_i) on X_i using OLS.
- In the log-linear model, an increase in X_i by one unit causes the expected value of ln(Y_i) to increase by β_1.
- An increase in ln(Y_i) by β_1 corresponds approximately to a $100 \times \beta_1\%$ increase in Y_i , since

$$\ln(Y_i + \Delta Y_i) \approx \ln(Y_i) + \frac{\Delta Y_i}{Y_i}.$$

2) Nonlinear functions of a single independent variable

- The *log-log* regression model is

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

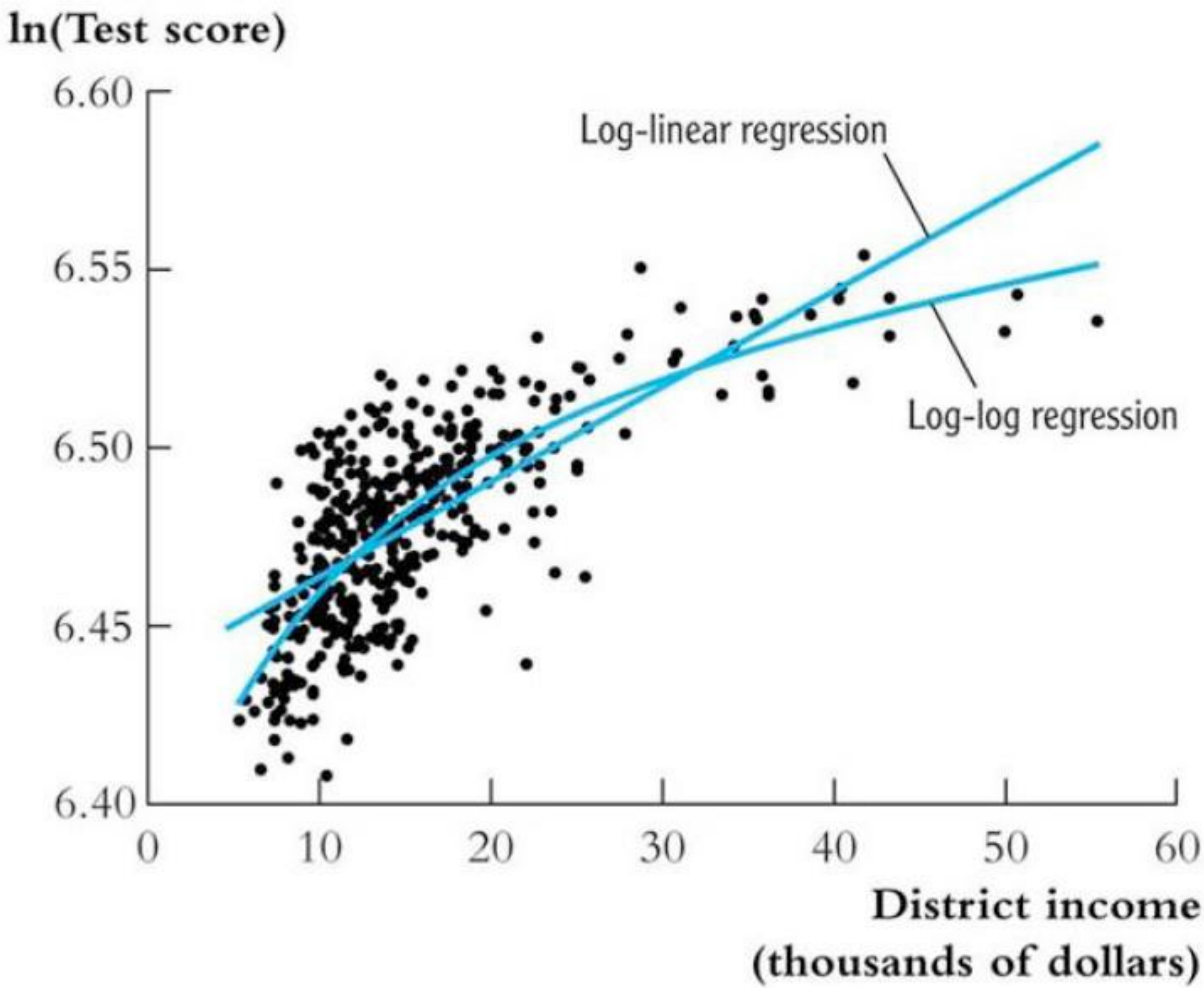
- Note that we can rewrite this equation as

$$Y_i = AX_i^{\beta_1} e^{u_i},$$

where $A = e^{\beta_0}$.

- We can estimate β_0 and β_1 by regressing $\ln(Y_i)$ on $\ln(X_i)$ using OLS.
- In the log-log model, an increase in Xi by 1% causes the expected value of $\ln(Y_i)$ to increase by approximately $0.01 \times \beta_1$.
- An increase in $\ln(Y_i)$ by $0.01 \times \beta_1$ corresponds approximately to a $\beta_1\%$ increase in Y_i .

2) Nonlinear functions of a single independent variable

- 

2) Nonlinear functions of a single independent variable

- In general, a good way to decide whether or to take logs of a dependent variable or regressor is to think about whether or not it makes more sense to think about absolute changes in the variable, or percentage changes.
- Economists typically model variables like income, consumption and stock prices in logs. This is because the magnitude of changes in these variables tends to be roughly proportional to the size of the variable.
- Variables that are expressed as rates, like the interest rate, the unemployment rate, the inflation rate, or the exchange rate, are more commonly left in levels. There are exceptions to this principle.

3) Interactions between independent variables

- The nonlinear models we have considered so far allow for the possibility that the change in Y_i caused by a change X_i may depend on the level of X_i .
- For instance, in our study of the relationship between test scores and average income, we found that an increase in income of $1000 had less effect on test scores when income was large. That is, we had diminishing returns to income. We were able to use polynomial, logarithmic, or other nonlinear regression equations to model this effect.
- An alternative possibility is that the change in Y_i caused by a change in X_i may depend on the level of some other variable. For instance, the effect of class size on test scores may depend on the level of English fluency in a classroom.
- We can include interaction terms in our regression equation to model these effects.

3) Interactions between independent variables

- Suppose our regression equation is

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

where the independent variables D_{1i} and D_{2i} are always equal to zero or one.

- For example, in our test score example:
  - D_{1i} equals zero for small class sizes (< 20%) and one for large class sizes ($\geq$20%), and
  - D_{2i} equals zero for classes with few English learners (< 10%) and one for classes with many English learners ($\geq$10%).
- In this model, the effect on Y_i of increasing D_{1i} from zero to one is equal to β_1. This effect does not depend on D_{2i}.
- In our example, this means we are assuming that a change in class size will have the same effect on test scores regardless of the students' level of English fluency.

3) Interactions between independent variables

- We might hypothesize that class size might matter more to English learners than to students already fluent in English. We can allow for this possibility by including an interaction variable in our regression equation. This interaction variable is simply the product of D_{1i} and D_{2i}:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i.$$

- Observe that

$$\begin{aligned}
E(Y_i | D_{1i} = 0, D_{2i} = 0) &= \beta_0 \\
E(Y_i | D_{1i} = 1, D_{2i} = 0) &= \beta_0 + \beta_1 \\
E(Y_i | D_{1i} = 0, D_{2i} = 1) &= \beta_0 + \beta_2 \\
E(Y_i | D_{1i} = 1, D_{2i} = 1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3.
\end{aligned}$$

3) Interactions between independent variables

- The effect on test scores of increasing class size from small to large in a class with few English learners is now

$$E(Y_i | D_{1i} = 1, D_{2i} = 0) - E(Y_i | D_{1i} = 0, D_{2i} = 0) = \beta_1 .$$

- The effect on test scores of increasing class size from small to large in a class with many English learners is

$$E(Y_i | D_{1i} = 1, D_{2i} = 1) - E(Y_i | D_{1i} = 0, D_{2i} = 1) = \beta_1 + \beta_3 .$$

- So, our model now allows for the effect of class size upon test scores to depend on English fluency.
- If we estimate this regression by OLS using the test score data, we find that $\hat{\beta}_3 = -3.5$, with a standard error of 3.1. The estimated coefficient is negative, as expected, but we cannot reject the null hypothesis that it is zero.

3) Interactions between independent variables

- We may also use interaction terms when one or both independent variables are continuous, rather than binary.

- When one variable is binary, and the other continuous, our regression equation is

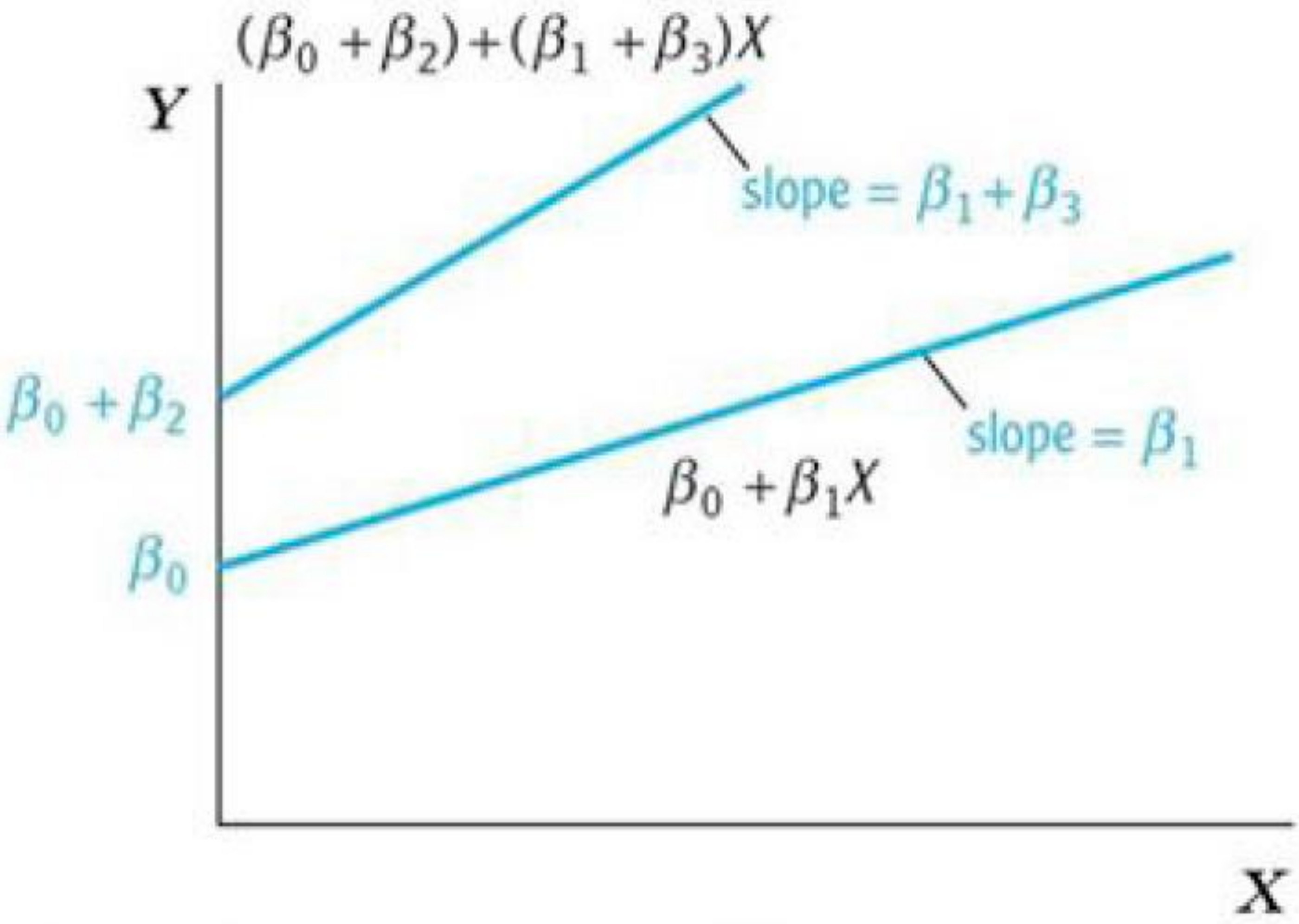
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i,$$

where X_i is continuous, meaning that it can be equal to any real number, and D_i is binary, meaning that it is always equal to zero or one.

- What is the effect upon Y_i of an increase in X_i by ΔX_i ?

$$\begin{aligned} E(\Delta Y_i | D_i = 0) &= \beta_1 \Delta X_i \\ E(\Delta Y_i | D_i = 1) &= (\beta_1 + \beta_3) \Delta X_i . \end{aligned}$$

3) Interactions between independent variables



3) Interactions between independent variables

- When both independent variables are continuous, our regression equation with an interaction effect is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i.$$

- What is the effect upon Y_i of an increase in X_{1i} by ΔX_{1i} ?

$$E(\Delta Y_i) = (\beta_1 + \beta_3 X_{2i}) \Delta X_{1i}.$$

- The effect of X_{1i} upon Y_i is now itself a linear function of X_{2i} .
- With the test score data, we estimate $\hat{\beta}_3 = 0.0012$, with a standard error of 0.019. Thus, we still cannot reject the null hypothesis that effect of class size upon test scores is not affected by English fluency levels.