BS2280 - Econometrics 1

Lecture 10 - Part 1: Nonlinear Models and Transformation of Variables II

Dr. Yichen Zhu

Structure of today's lecture

Intended Learning Outcomes

- Interpreting semi-log models
- Understanding dummy variables in log regression

Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 X^{\beta_2}$$

- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 logX$$
$$lnY = ln\beta_1 + \beta_2 lnX$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$dlnY = \beta_2 dlnX$$
$$\frac{dY}{Y} = \beta_2 \frac{dX}{X}$$

Derivative of Natural Log: $\frac{d}{dX} lnX = \frac{1}{X}$
$\frac{d}{dY} lnY = \frac{1}{Y}$
so $dlnX = \frac{dX}{X}$ *and* $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$100 \times \frac{dY}{Y} = \beta_2 100 \times \frac{dX}{X}$$
$$\%\Delta Y = \beta_2 \%\Delta X$$
$$\beta_2 = \frac{\%\Delta Y}{\%\Delta X} = elasticity$$

- We interpret as 1% change in X bring β_2% change in Y.

Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 X^{\beta_2}$$

- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 logX$$
$$lnY = ln\beta_1 + \beta_2 lnX$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

Derivative of Natural Log: $\frac{d}{dX} lnX = \frac{1}{X}$
$\frac{d}{dY} lnY = \frac{1}{Y}$
so $dlnX = \frac{dX}{X}$ and $dlnY = \frac{dY}{Y}$

$$dlnY = \beta_2 dlnX$$
$$\frac{dY}{Y} = \beta_2 \frac{dX}{X}$$

multiply by 100 for both sides

$$100 \times \frac{dY}{Y} = \beta_2 100 \times \frac{dX}{X}$$
$$\%\Delta Y = \beta_2 \%\Delta X$$
$$\beta_2 = \frac{\%\Delta Y}{\%\Delta X} = elasticity$$

- We interpret as 1% change in X bring β_2% change in Y.

Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 X^{\beta_2}$$

- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 logX$$
$$lnY = ln\beta_1 + \beta_2 lnX$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$dlnY = \beta_2 dlnX$$
$$\frac{dY}{Y} = \beta_2 \frac{dX}{X}$$

Derivative of Natural Log: $\frac{d}{dX} lnX = \frac{1}{X}$

$\frac{d}{dY} lnY = \frac{1}{Y}$

so $dlnX = \frac{dX}{X}$ *and* $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$100 \times \frac{dY}{Y} = \beta_2 100 \times \frac{dX}{X}$$
$$\%\Delta Y = \beta_2 \%\Delta X$$
$$\beta_2 = \frac{\%\Delta Y}{\%\Delta X} = elasticity$$

- We interpret as 1% change in X bring β_2% change in Y.

Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 X^{\beta_2}$$

- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 logX$$
$$lnY = ln\beta_1 + \beta_2 lnX$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

Derivative of Natural Log: $\frac{d}{dX} lnX = \frac{1}{X}$

$\frac{d}{dY} lnY = \frac{1}{Y}$

so $dlnX = \frac{dX}{X}$ and $dlnY = \frac{dY}{Y}$

$$dlnY = \beta_2 dlnX$$
$$\frac{dY}{Y} = \beta_2 \frac{dX}{X}$$

multiply by 100 for both sides

$$100 \times \frac{dY}{Y} = \beta_2 100 \times \frac{dX}{X}$$
$$\%\Delta Y = \beta_2 \% \Delta X$$
$$\beta_2 = \frac{\%\Delta Y}{\%\Delta X} = elasticity$$

- We interpret as 1% change in X bring β_2% change in Y.

Logarithmic Models: Example

$$FDHO_i = \beta_1 EXP_i^{\beta_2}$$

- After log transformation, We estimate the following model:

$$logFDHO_i = log\beta_1 + \beta_2 logEXP_i$$

- According to R output,

$$\widehat{logFDHO_i} = 0.70 + 0.66 logEXP_i$$

$log\beta_1 = 0.70$, so $\beta_1 = e^{0.701} = 2.02$

- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.

Logarithmic Models: Example

$$FDHO_i = \beta_1 EXP_i^{\beta_2}$$

- After log transformation, We estimate the following model:

$$logFDHO_i = log\beta_1 + \beta_2 logEXP_i$$

- According to R output,

$$\widehat{logFDHO_i} = 0.70 + 0.66 logEXP_i$$

$log\beta_1 = 0.70$, so $\beta_1 = e^{0.701} = 2.02$

- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.

Logarithmic Models: Example

$$FDHO_i = \beta_1 EXP_i^{\beta_2}$$

- After log transformation, We estimate the following model:

$$logFDHO_i = log\beta_1 + \beta_2 logEXP_i$$

- According to R output,

$$\widehat{logFDHO_i} = 0.70 + 0.66 logEXP_i$$

$log\beta_1 = 0.70$, so $\beta_1 = e^{0.701} = 2.02$

- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.

Benefits of Logarithmic Models

- Logarithmic transformation is commonly applied due to its advantages in the interpretation of the relationships
- Give a direct estimate of the elasticity
- Can improve the theoretical foundation of the model
- The logarithmic transformation improves the statistical specification of the model when the data is skewed
- It narrows the distribution, limiting the effect of outliers

Benefits of Logarithmic Models

- Logarithmic transformation is commonly applied due to its advantages in the interpretation of the relationships
- Give a direct estimate of the elasticity
- Can improve the theoretical foundation of the model
- The logarithmic transformation improves the statistical specification of the model when the data is skewed
- It narrows the distribution, limiting the effect of outliers

Benefits of Logarithmic Models

- Logarithmic transformation is commonly applied due to its advantages in the interpretation of the relationships
- Give a direct estimate of the elasticity
- Can improve the theoretical foundation of the model
- The logarithmic transformation improves the statistical specification of the model when the data is skewed
- It narrows the distribution, limiting the effect of outliers

Benefits of Logarithmic Models

- Logarithmic transformation is commonly applied due to its advantages in the interpretation of the relationships
- Give a direct estimate of the elasticity
- Can improve the theoretical foundation of the model
- The logarithmic transformation improves the statistical specification of the model when the data is skewed
- It narrows the distribution, limiting the effect of outliers

Benefits of Logarithmic Models

- Logarithmic transformation is commonly applied due to its advantages in the interpretation of the relationships
- Give a direct estimate of the elasticity
- Can improve the theoretical foundation of the model
- The logarithmic transformation improves the statistical specification of the model when the data is skewed
- It narrows the distribution, limiting the effect of outliers

Semi-Logarithmic Models

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 e^{\beta_2 X} && \text{Use: } logXZ = logX + logZ \\ &= log\beta_1 + loge^{\beta_2 X} && \text{Use: If } logX^n = nlogX \\ &= log\beta_1 + \beta_2 Xloge && \text{We know } loge = 1 \\ &= log\beta_1 + \beta_2 X \end{aligned}$$

- Y is measured in log, but X is not! X is measured in level.
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Semi-Logarithmic Models

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 e^{\beta_2 X} && \text{Use: } logXZ = logX + logZ \\ &= log\beta_1 + loge^{\beta_2 X} && \text{Use: If } logX^n = nlogX \\ &= log\beta_1 + \beta_2 Xloge && \text{We know } loge = 1 \\ &= log\beta_1 + \color{red}{\beta_2 X} \end{aligned}$$

- Y is measured in log, but X is not! X is measured in level.
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Semi-Logarithmic Models

- For the model
$$Y = \beta_1 e^{\beta_2 X}$$
- We use logarithmic transformation
$$\begin{aligned} logY &= log\beta_1 e^{\beta_2 X} && \text{Use: } logXZ = logX + logZ \\ &= log\beta_1 + loge^{\beta_2 X} && \text{Use: If } logX^n = nlogX \\ &= log\beta_1 + \beta_2 X loge && \text{We know } loge = 1 \\ &= log\beta_1 + \beta_2 X \end{aligned}$$
- Y is measured in log, but X is not! X is measured in level.
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Semi-Logarithmic Models

- For the model
$$Y = \beta_1 e^{\beta_2 X}$$
- We use logarithmic transformation
$$\begin{aligned} logY &= log\beta_1 e^{\beta_2 X} && \text{Use: } logXZ = logX + logZ \\ &= log\beta_1 + loge^{\beta_2 X} && \text{Use: If } logX^n = nlogX \\ &= log\beta_1 + \beta_2 Xloge && \text{We know } loge = 1 \\ &= log\beta_1 + {\color{red}\beta_2 X} \end{aligned}$$
- Y is measured in log, but X is not! X is measured in level.
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$\begin{aligned} dlnY &= \beta_2 dX \\ \frac{dY}{Y} &= \beta_2 dX \end{aligned}$$

Derivative of Natural Log: $\frac{d}{dY} lnY = \frac{1}{Y}$ so $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$\begin{aligned} 100 \times \frac{dY}{Y} &= 100 \times \beta_2 dX \\ \%\Delta Y &= 100\beta_2 \Delta X \\ \beta_2 &= \frac{\%\Delta Y}{100\Delta X} = semi-elasticity \end{aligned}$$

- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$\begin{aligned} dlnY &= \beta_2 dX \\ \frac{dY}{Y} &= \beta_2 dX \end{aligned}$$

Derivative of Natural Log: $\frac{d}{dY} lnY = \frac{1}{Y}$ so $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$\begin{aligned} 100 \times \frac{dY}{Y} &= 100 \times \beta_2 dX \\ \%\Delta Y &= 100\beta_2 \Delta X \\ \beta_2 &= \frac{\%\Delta Y}{100\Delta X} = semi - elasticity \end{aligned}$$

- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$\begin{aligned} dlnY &= \beta_2 dX \\ \frac{dY}{Y} &= \beta_2 dX \end{aligned}$$

Derivative of Natural Log: $\frac{d}{dY} lnY = \frac{1}{Y}$
so $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$\begin{aligned} 100 \times \frac{dY}{Y} &= 100 \times \beta_2 dX \\ \%\Delta Y &= 100\beta_2 \Delta X \\ \beta_2 &= \frac{\%\Delta Y}{100\Delta X} = semi - elasticity \end{aligned}$$

- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- To calculate the marginal impact of an increase in X on Y, we differentiate it.

$$\begin{aligned} dlnY &= \beta_2 dX \\ \frac{dY}{Y} &= \beta_2 dX \end{aligned}$$

Derivative of Natural Log: $\frac{d}{dY} lnY = \frac{1}{Y}$
so $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

$$\begin{aligned} 100 \times \frac{dY}{Y} &= 100 \times \beta_2 dX \\ \%\Delta Y &= 100\beta_2 \Delta X \\ \beta_2 &= \frac{\%\Delta Y}{100\Delta X} = semi - elasticity \end{aligned}$$

- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 X$$
$$lnY = ln\beta_1 + \beta_2 X$$

log is Natural Logarithm log_e or ln

- The interpretation of intercept:
- When $X = 0$, $Y = \beta_1 e^{\beta_2 \times 0} = \beta_1$ (note that $e^0 = 1$).
- β_1 is the value of Y when X is equal to zero.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- The interpretation of intercept:
- When $X = 0$, $Y = \beta_1 e^{\beta_2 \times 0} = \beta_1$ (note that $e^0 = 1$).
- β_1 is the value of Y when X is equal to zero.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- The interpretation of intercept:
- When $X = 0$, $Y = \beta_1 e^{\beta_2 \times 0} = \beta_1$ (note that $e^0 = 1$).
- β_1 is the value of Y when X is equal to zero.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} logY &= log\beta_1 + \beta_2 X \\ lnY &= ln\beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- The interpretation of intercept:
- When $X = 0$, $Y = \beta_1 e^{\beta_2 \times 0} = \beta_1$ (note that $e^0 = 1$).
- β_1 is the value of Y when X is equal to zero.

Semi-Logarithmic Models: Interpretations

- For the model

$$Y = \beta_1 e^{\beta_2 X}$$

- We use logarithmic transformation and get

$$\begin{aligned} log Y &= log \beta_1 + \beta_2 X \\ ln Y &= ln \beta_1 + \beta_2 X \end{aligned}$$

log is Natural Logarithm log_e or ln

- The interpretation of intercept:
- When $X = 0$, $Y = \beta_1 e^{\beta_2 \times 0} = \beta_1$ (note that $e^0 = 1$).
- β_1 is the value of Y when X is equal to zero.

Semi-Logarithmic Models: Example

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

```
Call:
lm(formula = lnEARN ~ S, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.83624    0.12894  14.241  < 2e-16 ***
S            0.06646    0.00853   7.792 3.88e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 0.5226 on 498 degrees of freedom
Multiple R-squared:  0.1087,	Adjusted R-squared:  0.1069
F-statistic: 60.71 on 1 and 498 DF,  p-value: 3.876e-14
```

- $$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

Semi-Logarithmic Models: Example

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

```
Call:
lm(formula = lnEARN ~ S, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.83624    0.12894  14.241  < 2e-16 ***
S            0.06646    0.00853   7.792 3.88e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 0.5226 on 498 degrees of freedom
Multiple R-squared:  0.1087,     Adjusted R-squared:  0.1069
F-statistic: 60.71 on 1 and 498 DF,  p-value: 3.876e-14
```

- $$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

Semi-Logarithmic Models: Example

$$\widehat{logEARN}_i = 1.83 + 0.066 S_i$$

- The estimate of the semi-elasticity is 0.066.
- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.
- An extra year of schooling increases hourly earnings by 100×0.066 %, which is 6.6%
- The intercept is $log\beta_1 = 1.83$, therefore, $\beta_1 = e^{1.83} = 6.27$.
- A person with no schooling would earn USD 6.27 per hour.

Semi-Logarithmic Models: Example

$$\widehat{logEARN_i} = 1.83 + 0.066S_i$$

- The estimate of the semi-elasticity is 0.066.
- We interpret as 1 unit change in X bring $100 \times \beta_2$% change in Y.
- An extra year of schooling increases hourly earnings by 100×0.066 %, which is 6.6%
- The intercept is $log\beta_1 = 1.83$, therefore, $\beta_1 = e^{1.83} = 6.27$.
- A person with no schooling would earn USD 6.27 per hour.

Semi-Logarithmic Models: Example

$$\widehat{log EARN_i} = 1.83 + 0.066 S_i$$

- The estimate of the semi-elasticity is 0.066.
- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.
- An extra year of schooling increases hourly earnings by 100×0.066 %, which is 6.6%
- The intercept is $log \beta_1 = 1.83$, therefore, $\beta_1 = e^{1.83} = 6.27$.
- A person with no schooling would earn USD 6.27 per hour.

Semi-Logarithmic Models: Example

$$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

- The estimate of the semi-elasticity is 0.066.
- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.
- An extra year of schooling increases hourly earnings by 100×0.066 %, which is 6.6%
- The intercept is $log \beta_1 = 1.83$, therefore, $\beta_1 = e^{1.83} = 6.27$.
- A person with no schooling would earn USD 6.27 per hour.

Semi-Logarithmic Models: Example

$$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

- The estimate of the semi-elasticity is 0.066.
- We interpret as 1 unit change in X bring $100 \times \beta_2\%$ change in Y.
- An extra year of schooling increases hourly earnings by 100×0.066 %, which is 6.6%
- The intercept is $log\beta_1 = 1.83$, therefore, $\beta_1 = e^{1.83} = 6.27$.
- A person with no schooling would earn USD 6.27 per hour.

Compare Linear and Nonlinear Model

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

- According to R output,

$$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

- $log\beta_1 = 1.83$, so $\beta_1 = e^{1.83} = 6.27$, $\beta_2 = 0.066$
- Therefore,

$$EARN_i = 6.27 e^{0.066 S_i}$$

- We can now compare the linear model and the nonlinear in the same diagram.

Compare Linear and Nonlinear Model

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

- According to R output,

$$\widehat{logEARN}_i = 1.83 + 0.066 S_i$$

- $log\beta_1 = 1.83$, so $\beta_1 = e^{1.83} = 6.27$, $\beta_2 = 0.066$
- Therefore,

$$EARN_i = 6.27 e^{0.066 S_i}$$

- We can now compare the linear model and the nonlinear in the same diagram.

Compare Linear and Nonlinear Model

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

- According to R output,

$$\widehat{logEARN}_i = 1.83 + 0.066 S_i$$

- $log\beta_1 = 1.83$, so $\beta_1 = e^{1.83} = 6.27$, $\beta_2 = 0.066$
- Therefore,

$$EARN_i = 6.27 e^{0.066 S_i}$$

- We can now compare the linear model and the nonlinear in the same diagram.

Compare Linear and Nonlinear Model

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

- According to R output,

$$\widehat{logEARN_i} = 1.83 + 0.066 S_i$$

- $log\beta_1 = 1.83$, so $\beta_1 = e^{1.83} = 6.27$, $\beta_2 = 0.066$
- Therefore,

$$EARN_i = 6.27 e^{0.066 S_i}$$

- We can now compare the linear model and the nonlinear in the same diagram.

Compare Linear and Nonlinear Model

$$EARN_i = \beta_1 e^{\beta_2 S_i}$$

- After log transformation, We estimate the following model:

$$logEARN_i = log\beta_1 + \beta_2 S_i$$

- According to R output,

$$\widehat{logEARN}_i = 1.83 + 0.066 S_i$$

- $log\beta_1 = 1.83$, so $\beta_1 = e^{1.83} = 6.27$, $\beta_2 = 0.066$
- Therefore,

$$EARN_i = 6.27 e^{0.066 S_i}$$

- We can now compare the linear model and the nonlinear in the same diagram.

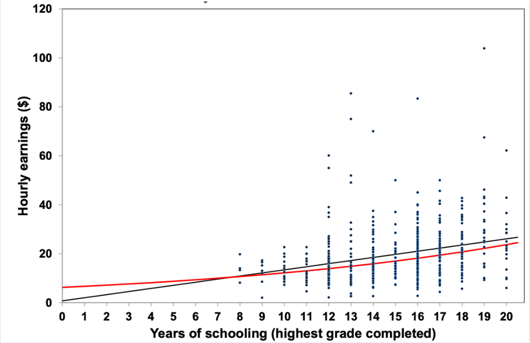
Compare Linear and Nonlinear Model

- Linear model:

$$\widehat{EARN}_i = 0.76 + 1.26S_i$$

- Nonlinear model:

$$EARN_i = 6.27e^{0.066S_i}$$



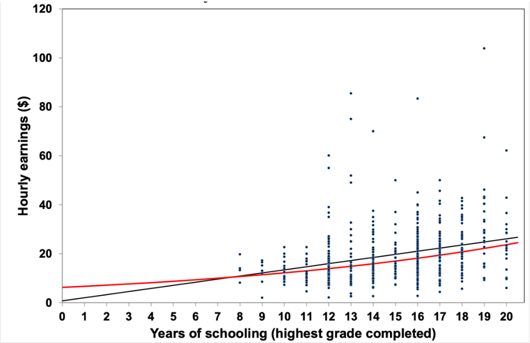
Compare Linear and Nonlinear Model

- Linear model:

$$\widehat{EARN}_i = 0.76 + 1.26 S_i$$

- Nonlinear model:

$$EARN_i = 6.27 e^{0.066 S_i}$$



When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

When We Use Log?

- Rule of thumb
- What types of variables are often used in log form?
  1. Monetary amounts that must be positive
  2. Very large variables, such as population
- What types of variables are often used in level form?
  1. Variables that are a proportion or percent

Summary

Model	**Dependent Variable**	**Independent Variable**	**Interpretation**
$Y = \beta_1 + \beta_2 X$ Level-Level Model	Y	X	1 unit change in X bring β_2 units change in Y
$logY = \beta_1 + \beta_2 logX$ Log-Log Model	$logY$	$logX$	1 % change in X bring β_2 % change in Y
$logY = \beta_1 + \beta_2 X$ Log-Level Model	$logY$	X	1 unit change in X bring $100 \times \beta_2$ % change in Y

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Dummy Variables in Log Regression

- We often add dummies to a logarithmic regression. For example,

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- **DO NOT LOG THE DUMMY!!!!**
- With a dummy, and a logarithmic dependent variable, we are actually estimating a Log-Level model (or semi-log model)
- But a dummy is not a continuous variable, therefore we cannot measure infinitely small marginal changes!
- How do we interpret dummy variables in a log regression?
- It will depend on the value of the coefficient.

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$

- **Example**:

$$\widehat{logEARN_i} = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$
- **Example**:

$$\widehat{logEARN_i} = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$

- **Example**:

$$\widehat{logEARN_i} = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Small Coefficient

- For small coefficients, we can interpret the coefficient like any other coefficient in a Log-Level model (or semi-log model)
- Small coefficient range: $-0.3 \leq \hat{\delta} \leq 0.3$
- Interpretation: If the dummy variable is equal to 1, then Y increases by $100 \times \hat{\delta}\%$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: For female, hourly earnings increase by $100 \times \hat{\delta}\%$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN_i} = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable $FEMALE$ is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable $FEMALE$ is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient

- For large coefficients, the interpretations are relatively complicated
- Large coefficient range: $\hat{\delta} < -0.3$ and $\hat{\delta} > 0.3$
- Interpretation: The Y differences between dummy is 1 category and base category are $100 \times (e^{\hat{\delta}} - 1)$

- **Example**:

$$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$$

- Dummy variable *FEMALE* is 1 for female and 0 for male
- Interpretation: The hourly earnings differences between female and male are $100 \times (e^{\hat{\delta}} - 1)$ \$/hour

Large Coefficient: Derivations

	$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta} FEMALE_i$	
Male	$FEMALE = 0$	$\widehat{logEARN}_i = \hat{\beta}_1$
Female	$FEMALE = 1$	$\widehat{logEARN}_i = \hat{\beta}_1 + \hat{\delta}$

- Therefore:

$$\hat{\delta} = logEARN_{i,female} - logEARN_{i,male}$$

Small and Large Coefficient

Coefficient	%	Exact	Coefficient	%	Exact
0.5	50	64.9	0	0	0
0.4	40	49.2	-0.05	-5	-4.9
0.35	35	41.9	-0.1	-10	-9.5
0.3	30	35.0	-0.15	-15	-13.9
0.25	25	28.4	-0.2	-20	-18.1
0.2	20	22.1	-0.25	-25	-22.1
0.15	15	16.2	-0.3	-30	-25.9
0.1	10	10.5	-0.35	-35	-29.5
0.05	5	5.1	-0.4	-40	-33.0
0	0	0	-0.5	-50	-39.3

A Further Complication

- $\hat{\delta}$ can be positive or negative
- The size of the impact will depend on the sign, therefore the selection of the reference group is important!
- See table before: E.g. based on the reference category, coefficient may be -0.4 or +0.4
- The exact impact may be -33% or +49.2%!
- Using the quick method will give us a value in the middle, independent of the reference category

A Further Complication

- $\hat{\delta}$ can be positive or negative
- The size of the impact will depend on the sign, therefore the selection of the reference group is important!
- See table before: E.g. based on the reference category, coefficient may be -0.4 or +0.4
- The exact impact may be -33% or +49.2%!
- Using the quick method will give us a value in the middle, independent of the reference category

A Further Complication

- $\hat{\delta}$ can be positive or negative
- The size of the impact will depend on the sign, therefore the selection of the reference group is important!
- See table before: E.g. based on the reference category, coefficient may be -0.4 or +0.4
- The exact impact may be -33% or +49.2%!
- Using the quick method will give us a value in the middle, independent of the reference category

A Further Complication

- $\hat{\delta}$ can be positive or negative
- The size of the impact will depend on the sign, therefore the selection of the reference group is important!
- See table before: E.g. based on the reference category, coefficient may be -0.4 or +0.4
- The exact impact may be -33% or +49.2%!
- Using the quick method will give us a value in the middle, independent of the reference category

A Further Complication

- $\hat{\delta}$ can be positive or negative
- The size of the impact will depend on the sign, therefore the selection of the reference group is important!
- See table before: E.g. based on the reference category, coefficient may be -0.4 or +0.4
- The exact impact may be -33% or +49.2%!
- Using the quick method will give us a value in the middle, independent of the reference category