BS2280 - Econometrics 1

Lecture 9 - Part 2: Nonlinear Models and Transformation of Variables I

Dr. Yichen Zhu

Structure of today's lecture

1. Models which are nonlinear in variables: Hyperbolic Model
2. Models which are nonlinear in parameters: Logarithmic Models

Intended Learning Outcomes

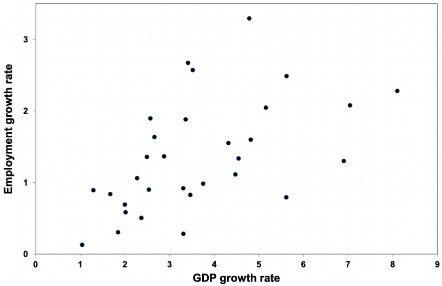
- Understanding methods how models that are non-linear in variables and parameters can be estimated with OLS
- Interpreting the coefficients of
  - Hyperbolic models
  - Logarithmic models

Background

- We will begin with an example of a simple model that can be linearised by a cosmetic transformation.
- The table and plot show the average annual rates of growth of employment and GDP for 31 OECD countries

Average annual percentage growth rates

	Employment	GDP		Employment	GDP
Australia	2.57	3.52	Korea	1.11	4.48
Austria	1.64	2.66	Luxembourg	1.34	4.55
Belgium	1.06	2.27	Mexico	1.88	3.36
Canada	1.90	2.57	Netherlands	0.51	2.37
Czech Republic	0.79	5.62	New Zealand	2.67	3.41
Denmark	0.58	2.02	Norway	1.36	2.49
Estonia	2.28	8.10	Poland	2.05	5.16
Finland	0.98	3.75	Portugal	0.13	1.04
France	0.69	2.00	Slovak Republic	2.08	7.04
Germany	0.84	1.67	Slovenia	1.60	4.82
Greece	1.55	4.32	Sweden	0.83	3.47
Hungary	0.28	3.31	Switzerland	0.90	2.54
Iceland	2.49	5.62	Turkey	1.30	6.90
Israel	3.29	4.79	United Kingdom	0.92	3.31
Italy	0.89	1.29	United States	1.36	2.88
Japan	0.31	1.85			



- We would expect nonlinear relationship!

Review: Nonlinear in Variables

- **Linear in variables**: The variables included on the right side of the regression are exactly as defined, rather than as functions
- Linear in variables:

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in variables:

$$EARNINGS_i = \beta_1 + \beta_2 S_i^2 + \beta_3 \sqrt{EXP_i} + u_i$$

Models which are nonlinear in variables: Hyperbolic Model

- This model is nonlinear in variables, because variable g is expressed as $\frac{1}{g}$

$$e_i = \beta_1 + \frac{\beta_2}{g_i} + u_i$$

- We can rewrite the model so that is linear in variables as well as parameters
- Define $z = \frac{1}{g}$
- Therefore

$$e_i = \beta_1 + \beta_2 z_i + u_i$$

- Data table with computed values of z

Average annual percentage growth rates

	e	g	z		e	g	z
Australia	2.57	3.52	0.2841	Korea	1.11	4.48	0.2235
Austria	1.64	2.66	0.3757	Luxembourg	1.34	4.55	0.2199
Belgium	1.06	2.27	0.4401	Mexico	1.88	3.36	0.2976
Canada	1.90	2.57	0.3891	Netherlands	0.51	2.37	0.4221
Czech Republic	0.79	5.62	0.1781	New Zealand	2.67	3.41	0.2929
Denmark	0.58	2.02	0.4961	Norway	1.36	2.49	0.4013
Estonia	2.28	8.10	0.1234	Poland	2.05	5.16	0.1938
Finland	0.98	3.75	0.2664	Portugal	0.13	1.04	0.9603
France	0.69	2.00	0.5004	Slovak Republic	2.08	7.04	0.1420
Germany	0.84	1.67	0.5980	Slovenia	1.60	4.82	0.2075
Greece	1.55	4.32	0.2315	Sweden	0.83	3.47	0.2885
Hungary	0.28	3.31	0.3021	Switzerland	0.90	2.54	0.3941
Iceland	2.49	5.62	0.1779	Turkey	1.30	6.90	0.1449
Israel	3.29	4.79	0.2089	United Kingdom	0.92	3.31	0.3024
Italy	0.89	1.29	0.7723	United States	1.36	2.88	0.3476
Japan	0.31	1.85	0.5417				

Models which are nonlinear in variables: Hyperbolic Model

- This model is nonlinear in variables, because variable g is expressed as $\frac{1}{g}$

$$e_i = \beta_1 + \frac{\beta_2}{g_i} + u_i$$

- We can rewrite the model so that is linear in variables as well as parameters
- Define $z = \frac{1}{g}$
- Therefore

$$e_i = \beta_1 + \beta_2 z_i + u_i$$

- Data table with computed values of z

Average annual percentage growth rates

	e	g	z		e	g	z
Australia	2.57	3.52	0.2841	Korea	1.11	4.48	0.2235
Austria	1.64	2.66	0.3757	Luxembourg	1.34	4.55	0.2199
Belgium	1.06	2.27	0.4401	Mexico	1.88	3.36	0.2976
Canada	1.90	2.57	0.3891	Netherlands	0.51	2.37	0.4221
Czech Republic	0.79	5.62	0.1781	New Zealand	2.67	3.41	0.2929
Denmark	0.58	2.02	0.4961	Norway	1.36	2.49	0.4013
Estonia	2.28	8.10	0.1234	Poland	2.05	5.16	0.1938
Finland	0.98	3.75	0.2664	Portugal	0.13	1.04	0.9603
France	0.69	2.00	0.5004	Slovak Republic	2.08	7.04	0.1420
Germany	0.84	1.67	0.5980	Slovenia	1.60	4.82	0.2075
Greece	1.55	4.32	0.2315	Sweden	0.83	3.47	0.2885
Hungary	0.28	3.31	0.3021	Switzerland	0.90	2.54	0.3941
Iceland	2.49	5.62	0.1779	Turkey	1.30	6.90	0.1449
Israel	3.29	4.79	0.2089	United Kingdom	0.92	3.31	0.3024
Italy	0.89	1.29	0.7723	United States	1.36	2.88	0.3476
Japan	0.31	1.85	0.5417				

Models which are nonlinear in variables: Hyperbolic Model

- This model is nonlinear in variables, because variable g is expressed as $\frac{1}{g}$

$$e_i = \beta_1 + \frac{\beta_2}{g_i} + u_i$$

- We can rewrite the model so that is linear in variables as well as parameters
- Define $z = \frac{1}{g}$
- Therefore

$$e_i = \beta_1 + \beta_2 z_i + u_i$$

- Data table with computed values of z

Average annual percentage growth rates

	e	g	z		e	g	z
Australia	2.57	3.52	0.2841	Korea	1.11	4.48	0.2235
Austria	1.64	2.66	0.3757	Luxembourg	1.34	4.55	0.2199
Belgium	1.06	2.27	0.4401	Mexico	1.88	3.36	0.2976
Canada	1.90	2.57	0.3891	Netherlands	0.51	2.37	0.4221
Czech Republic	0.79	5.62	0.1781	New Zealand	2.67	3.41	0.2929
Denmark	0.58	2.02	0.4961	Norway	1.36	2.49	0.4013
Estonia	2.28	8.10	0.1234	Poland	2.05	5.16	0.1938
Finland	0.98	3.75	0.2664	Portugal	0.13	1.04	0.9603
France	0.69	2.00	0.5004	Slovak Republic	2.08	7.04	0.1420
Germany	0.84	1.67	0.5980	Slovenia	1.60	4.82	0.2075
Greece	1.55	4.32	0.2315	Sweden	0.83	3.47	0.2885
Hungary	0.28	3.31	0.3021	Switzerland	0.90	2.54	0.3941
Iceland	2.49	5.62	0.1779	Turkey	1.30	6.90	0.1449
Israel	3.29	4.79	0.2089	United Kingdom	0.92	3.31	0.3024
Italy	0.89	1.29	0.7723	United States	1.36	2.88	0.3476
Japan	0.31	1.85	0.5417				

Models which are nonlinear in variables: Hyperbolic Model

- This model is nonlinear in variables, because variable g is expressed as $\frac{1}{g}$

$$e_i = \beta_1 + \frac{\beta_2}{g_i} + u_i$$

- We can rewrite the model so that is linear in variables as well as parameters
- Define $z = \frac{1}{g}$
- Therefore

$$e_i = \beta_1 + \beta_2 z_i + u_i$$

- Data table with computed values of z

Average annual percentage growth rates

	e	g	z		e	g	z
Australia	2.57	3.52	0.2841	Korea	1.11	4.48	0.2235
Austria	1.64	2.66	0.3757	Luxembourg	1.34	4.55	0.2199
Belgium	1.06	2.27	0.4401	Mexico	1.88	3.36	0.2976
Canada	1.90	2.57	0.3891	Netherlands	0.51	2.37	0.4221
Czech Republic	0.79	5.62	0.1781	New Zealand	2.67	3.41	0.2929
Denmark	0.58	2.02	0.4961	Norway	1.36	2.49	0.4013
Estonia	2.28	8.10	0.1234	Poland	2.05	5.16	0.1938
Finland	0.98	3.75	0.2664	Portugal	0.13	1.04	0.9603
France	0.69	2.00	0.5004	Slovak Republic	2.08	7.04	0.1420
Germany	0.84	1.67	0.5980	Slovenia	1.60	4.82	0.2075
Greece	1.55	4.32	0.2315	Sweden	0.83	3.47	0.2885
Hungary	0.28	3.31	0.3021	Switzerland	0.90	2.54	0.3941
Iceland	2.49	5.62	0.1779	Turkey	1.30	6.90	0.1449
Israel	3.29	4.79	0.2089	United Kingdom	0.92	3.31	0.3024
Italy	0.89	1.29	0.7723	United States	1.36	2.88	0.3476
Japan	0.31	1.85	0.5417				

Models which are nonlinear in variables: Hyperbolic Model

- This model is nonlinear in variables, because variable g is expressed as $\frac{1}{g}$

$$e_i = \beta_1 + \frac{\beta_2}{g_i} + u_i$$

- We can rewrite the model so that is linear in variables as well as parameters
- Define $z = \frac{1}{g}$
- Therefore

$$e_i = \beta_1 + \beta_2 z_i + u_i$$

- Data table with computed values of z

Average annual percentage growth rates

	e	g	z		e	g	z
Australia	2.57	3.52	0.2841	Korea	1.11	4.48	0.2235
Austria	1.64	2.66	0.3757	Luxembourg	1.34	4.55	0.2199
Belgium	1.06	2.27	0.4401	Mexico	1.88	3.36	0.2976
Canada	1.90	2.57	0.3891	Netherlands	0.51	2.37	0.4221
Czech Republic	0.79	5.62	0.1781	New Zealand	2.67	3.41	0.2929
Denmark	0.58	2.02	0.4961	Norway	1.36	2.49	0.4013
Estonia	2.28	8.10	0.1234	Poland	2.05	5.16	0.1938
Finland	0.98	3.75	0.2664	Portugal	0.13	1.04	0.9603
France	0.69	2.00	0.5004	Slovak Republic	2.08	7.04	0.1420
Germany	0.84	1.67	0.5980	Slovenia	1.60	4.82	0.2075
Greece	1.55	4.32	0.2315	Sweden	0.83	3.47	0.2885
Hungary	0.28	3.31	0.3021	Switzerland	0.90	2.54	0.3941
Iceland	2.49	5.62	0.1779	Turkey	1.30	6.90	0.1449
Israel	3.29	4.79	0.2089	United Kingdom	0.92	3.31	0.3024
Italy	0.89	1.29	0.7723	United States	1.36	2.88	0.3476
Japan	0.31	1.85	0.5417				

Hyperbolic Model

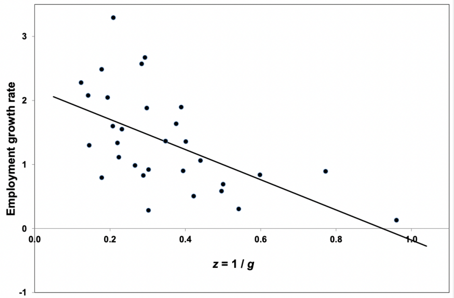
- Output regression of e on z.

```
Call:
lm(formula = e ~ z, data = oecd_exercises)

Residuals:
     Min       1Q   Median       3Q      Max
-1.18370 -0.53588 -0.07868  0.37322  1.60711

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1731     0.2495   8.710 1.37e-09 ***
z            -2.3481     0.6367  -3.688 0.000926 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6522 on 29 degrees of freedom
Multiple R-squared:  0.3193,    Adjusted R-squared:  0.2958
F-statistic:  13.6 on 1 and 29 DF,  p-value: 0.0009261
```

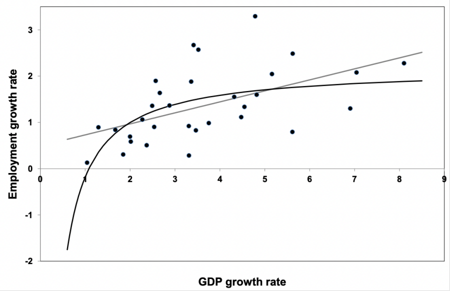


- The figure shows the transformed data and the regression line for the regression of e on z.
- Therefore

$$\hat{e}_i = 2.17 - 2.35 z_i$$

Hyperbolic Model

- However, we define $z = \frac{1}{g}$
- So, substituting $\frac{1}{g}$ for z, we obtain the nonlinear relationship between e and g.
- The figure shows this relationship plotted in the original diagram.



- Therefore

$$\hat{e}_i = 2.17 - 2.35 z_i = 2.17 - \frac{2.35}{g_i}$$

Hyperbolic Model: Interpretations

$$\hat{e}_i = 2.17 - 2.35 z_i = 2.17 - \frac{2.35}{g_i}$$

- **Question**: How can we interpret this model? What is the marginal impact of an increase in g on e?
- To calculate the marginal impact of an increase in g on e, we have to take the first derivative.
  1. Derive

     $$\frac{d\hat{e}_i}{dg_i} =$$

     (Hint: Make use of the power rule, derivative of $\frac{1}{x}$ is $-\frac{1}{x^2}$)
  2. What is the marginal impact of g on e when g is 1% and 3%?
- The marginal effect will not be constant!

Hyperbolic Model: Interpretations

$$\hat{e}_i = 2.17 - 2.35z_i = 2.17 - \frac{2.35}{g_i}$$

- **Question**: How can we interpret this model? What is the marginal impact of an increase in g on e?
- To calculate the marginal impact of an increase in g on e, we have to take the first derivative.
  1. Derive
  $$\frac{d\hat{e}_i}{dg_i} =$$
  (Hint: Make use of the power rule, derivative of $\frac{1}{x}$ is $-\frac{1}{x^2}$)
  2. What is the marginal impact of g on e when g is 1% and 3%?
- The marginal effect will not be constant!

Hyperbolic Model: Interpretations

$$\hat{e}_i = 2.17 - 2.35 z_i = 2.17 - \frac{2.35}{g_i}$$

- **Question**: How can we interpret this model? What is the marginal impact of an increase in g on e?
- To calculate the marginal impact of an increase in g on e, we have to take the first derivative.
  1. Derive

     $$\frac{d\hat{e}_i}{dg_i} =$$

     (Hint: Make use of the power rule, derivative of $\frac{1}{x}$ is $-\frac{1}{x^2}$)
  2. What is the marginal impact of g on e when g is 1% and 3%?
- The marginal effect will not be constant!

Hyperbolic Model: Interpretations

$$\hat{e}_i = 2.17 - 2.35 z_i = 2.17 - \frac{2.35}{g_i}$$

- **Question**: How can we interpret this model? What is the marginal impact of an increase in g on e?
- To calculate the marginal impact of an increase in g on e, we have to take the first derivative.
  1. Derive
     $$\frac{d\hat{e}_i}{dg_i} =$$
     (Hint: Make use of the power rule, derivative of $\frac{1}{x}$ is $-\frac{1}{x^2}$)
  2. What is the marginal impact of g on e when g is 1% and 3%?
- The marginal effect will not be constant!

Hyperbolic Model: Interpretations

$$\hat{e}_i = 2.17 - 2.35 z_i = 2.17 - \frac{2.35}{g_i}$$

- **Question**: How can we interpret this model? What is the marginal impact of an increase in g on e?
- To calculate the marginal impact of an increase in g on e, we have to take the first derivative.
  1. Derive
  $$\frac{d\hat{e}_i}{dg_i} =$$
  (Hint: Make use of the power rule, derivative of $\frac{1}{x}$ is $-\frac{1}{x^2}$)
  2. What is the marginal impact of g on e when g is 1% and 3%?
- The marginal effect will not be constant!

Background

- You will frequently come across data where a model with nonlinear parameter leads to a better fit of the model
- Fortunately, some of these models can be estimated with OLS nevertheless
- The natural Logarithm can do the magic!
- A Logarithmic transformation has also many positive externalities.

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Review: Nonlinear in Parameters

- **Linear in parameters:** The parameters $\beta_2, \beta_3, ..., \beta_k$ are multiplied with the X variables.
- **Assumption 1.** Model is linear in parameters and correctly specified:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

- **Example.**
- Linear in parameters (this model is also linear in variables):

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Nonlinear in parameters:

$$EARNINGS_i = \beta_1 + S_i^{\beta_2} + \beta_2 \beta_3 EXP_i + u_i$$

- The parameter is not just a simple multiplication with S_i or EXP_i anymore!!!

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^X = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^X = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^X = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Models which are nonlinear in parameters: Logarithmic Models

- Assume we have the following model:

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

- Clearly non-linear in parameters and OLS does not work with model
- However, We can make equation linear in parameters!
- We use logarithmic transformation

- Review: Basic rules of Logs
  - $logXZ = logX + logZ$
  - $log\frac{X}{Z} = logX - logZ$
  - $logX^n = nlogX$
  - $loge^x = X$ (In Econometrics, log is Natural Logarithm log_e or ln)

Logarithmic Models

- For the model

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

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 X^{\beta_2} \\ &= log\beta_1 + logX^{\beta_2} \\ &= log\beta_1 + \beta_2 logX \end{aligned}$$

Use: $logXZ = logX + logZ$

Use: If $logX^n = nlogX$

- We can simplify by writing:

$$Y' = \beta_1' + \beta_2 X'$$

- where $Y' = logY$, $\beta_1' = log\beta_1$, $X' = logX$
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Logarithmic Models

- For the model

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

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 X^{\beta_2} \\ &= log\beta_1 + logX^{\beta_2} \\ &= log\beta_1 + \beta_2 logX \end{aligned}$$

Use: $logXZ = logX + logZ$

Use: If $logX^n = nlogX$

- We can simplify by writing:

$$Y' = \beta_1' + \beta_2 X'$$

- where $Y' = logY$, $\beta_1' = log\beta_1$, $X' = logX$
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Logarithmic Models

- For the model

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

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 X^{\beta_2} \\ &= log\beta_1 + logX^{\beta_2} \\ &= log\beta_1 + \beta_2 logX \end{aligned}$$

Use: $logXZ = logX + logZ$

Use: If $logX^n = nlogX$

- We can simplify by writing:

$$Y' = \beta_1' + \beta_2 X'$$

- where $Y' = logY$, $\beta_1' = log\beta_1$, $X' = logX$
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Logarithmic Models

- For the model
$$Y = \beta_1 X^{\beta_2}$$
- We use logarithmic transformation
$$\begin{aligned} logY &= log\beta_1 X^{\beta_2} \\ &= log\beta_1 + logX^{\beta_2} \\ &= log\beta_1 + \beta_2 logX \end{aligned}$$
Use: $logXZ = logX + logZ$
Use: If $logX^n = nlogX$
- We can simplify by writing:
$$Y' = \beta_1' + \beta_2 X'$$
- where $Y' = logY$, $\beta_1' = log\beta_1$, $X' = logX$
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Logarithmic Models

- For the model

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

- We use logarithmic transformation

$$\begin{aligned} logY &= log\beta_1 X^{\beta_2} \\ &= log\beta_1 + logX^{\beta_2} \\ &= log\beta_1 + \beta_2 logX \end{aligned}$$

Use: $logXZ = logX + logZ$

Use: If $logX^n = nlogX$

- We can simplify by writing:

$$Y' = \beta_1' + \beta_2 X'$$

- where $Y' = logY$, $\beta_1' = log\beta_1$, $X' = logX$
- **Question**: How can we interpret β_2? What is the marginal impact of an increase in X on Y?

Logarithmic Models: Interpretations

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

- For the model
- We use logarithmic transformation and get

$$logY = log\beta_1 + \beta_2 logX$$

log is Natural Logarithm log_e or ln

$$lnY = ln\beta_1 + \beta_2 lnX$$

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

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

$$dlnY = \beta_2 dlnX$$

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

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

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

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 $\beta_2\%$ change in Y.

Logarithmic Models: Interpretations

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

- For the model
- 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

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

- For the model
- 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: $\frac{d}{dX} lnX = \frac{1}{X}$

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

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

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

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

- For the model
- 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: $\frac{d}{dX} lnX = \frac{1}{X}$

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

$\frac{d}{dY} lnY = \frac{1}{Y}$
so $dlnX = \frac{dX}{X}$ *and* $dlnY = \frac{dY}{Y}$

multiply by 100 for both sides

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

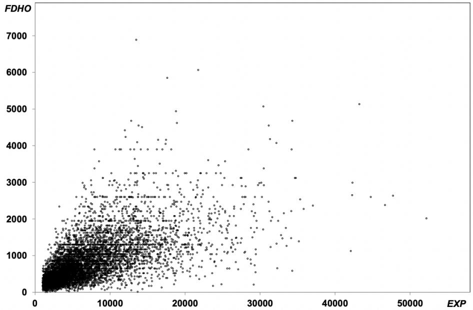
- We interpret as 1% change in X bring β_2% change in Y.

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

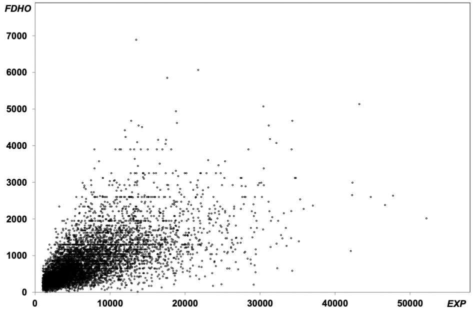
Logarithmic Models: Example

- We use data from the US Consumer Expenditure Survey data.
- We regress *FDHO*, food eaten at home, on *EXP*, total annual household expenditure
- Both are measured in dollars for a sample of 869 households in the United States
- Firstly, we present original data in USD



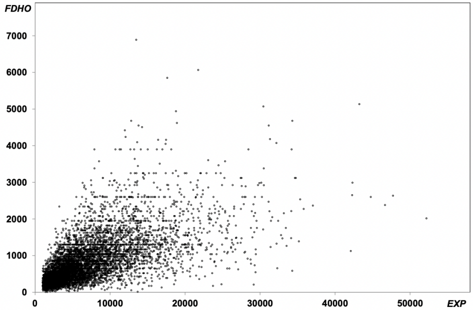
Logarithmic Models: Example

- We use data from the US Consumer Expenditure Survey data.
- We regress *FDHO*, food eaten at home, on *EXP*, total annual household expenditure
- Both are measured in dollars for a sample of 869 households in the United States
- Firstly, we present original data in USD



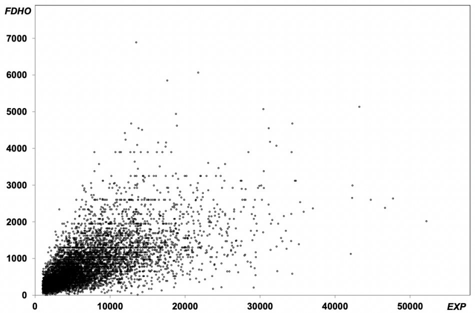
Logarithmic Models: Example

- We use data from the US Consumer Expenditure Survey data.
- We regress *FDHO*, food eaten at home, on *EXP*, total annual household expenditure
- Both are measured in dollars for a sample of 869 households in the United States
- Firstly, we present original data in USD



Logarithmic Models: Example

- We use data from the US Consumer Expenditure Survey data.
- We regress *FDHO*, food eaten at home, on *EXP*, total annual household expenditure
- Both are measured in dollars for a sample of 869 households in the United States
- Firstly, we present original data in USD



Logarithmic Models: Example

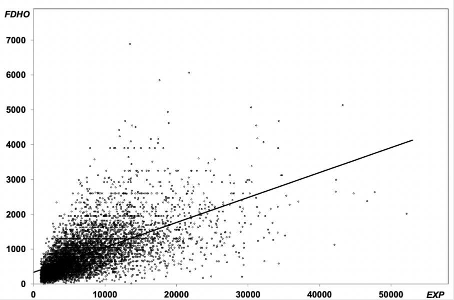
$$FDHO_i = \beta_1 + \beta_2 EXP_i + u_i$$

- We estimate the following model:

```
Call:
lm(formula = FDHO ~ EXP, data = ces2013)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 369.441758  10.657178   34.67   <2e-16 ***
EXP           0.062710   0.001071   58.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 532.4 on 6332 degrees of freedom
Multiple R-squared:  0.3514,	Adjusted R-squared:  0.3513
F-statistic:  3431 on 1 and 6332 DF,  p-value: < 2.2e-16
```



$$\widehat{FDHO}_i = 369.44 + 0.063 EXP_i$$

- 6.3 cents out of each dollar of expenditure is spent on food at home
- USD 369 would be spent on food at home if total expenditure were zero. Obviously this is impossible.

Logarithmic Models: Example

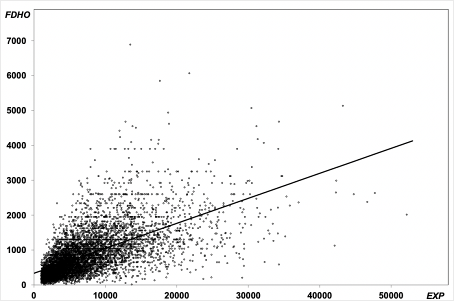
$$FDHO_i = \beta_1 + \beta_2 EXP_i + u_i$$

- We estimate the following model:

```
Call:
lm(formula = FDHO ~ EXP, data = ces2013)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 369.441758  10.657178   34.67   <2e-16 ***
EXP           0.062710   0.001071   58.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 532.4 on 6332 degrees of freedom
Multiple R-squared:  0.3514,  Adjusted R-squared:  0.3513
F-statistic:  3431 on 1 and 6332 DF,  p-value: < 2.2e-16
```



$$\widehat{FDHO_i} = 369.44 + 0.063 EXP_i$$

- 6.3 cents out of each dollar of expenditure is spent on food at home
- USD 369 would be spent on food at home if total expenditure were zero. Obviously this is impossible.

Logarithmic Models: Example

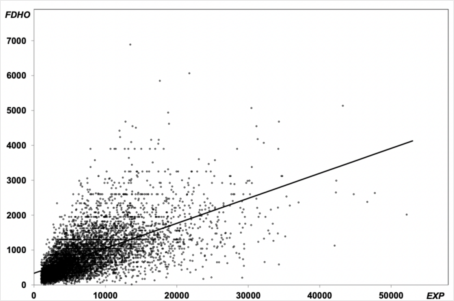
$$FDHO_i = \beta_1 + \beta_2 EXP_i + u_i$$

- We estimate the following model:

```
Call:
lm(formula = FDHO ~ EXP, data = ces2013)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 369.441758  10.657178   34.67   <2e-16 ***
EXP           0.062710   0.001071   58.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 532.4 on 6332 degrees of freedom
Multiple R-squared:  0.3514,  Adjusted R-squared:  0.3513
F-statistic:  3431 on 1 and 6332 DF,  p-value: < 2.2e-16
```

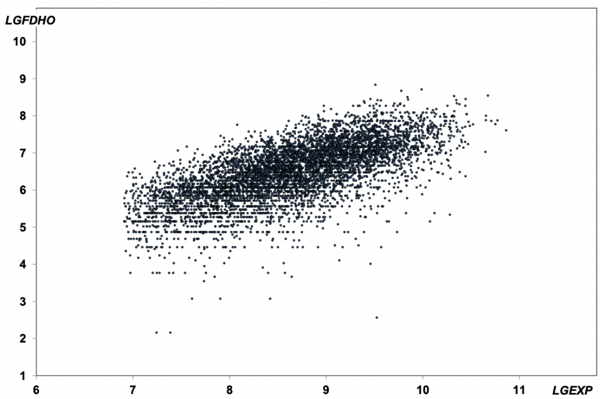


$$\widehat{FDHO}_i = 369.44 + 0.063 EXP_i$$

- 6.3 cents out of each dollar of expenditure is spent on food at home
- USD 369 would be spent on food at home if total expenditure were zero. Obviously this is impossible.

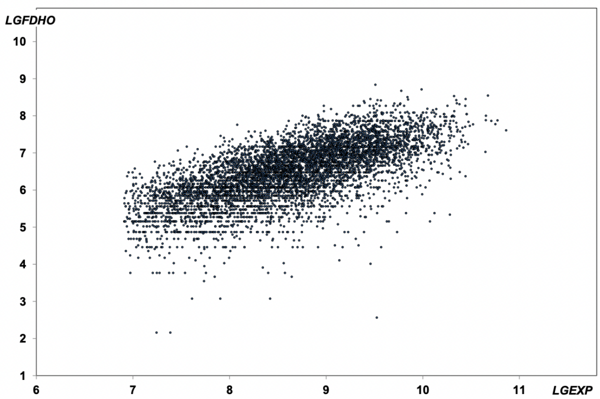
Logarithmic Models: Example

- We will fit now a constant elasticity function using the same data
- We begin with taking the natural Log of *FDHO* and *EXP* and plot the data below



Logarithmic Models: Example

- We will fit now a constant elasticity function using the same data
- We begin with taking the natural Log of *FDHO* and *EXP* and plot the data below



Logarithmic Models: Example

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

- After log transformation, We estimate the following model:

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

```
ces2013$lnFDHO <- log(ces2013$FDHO) # log transformation
sum(is.infinite(ces2013$lnFDHO)) # Are we losing obs because of log?
ces2013$lnEXP <- log(ces2013$EXP)
sum(is.infinite(ces2013$lnEXP))
```

```
Call:
lm(formula = lnFDHO ~ lnEXP, data = ces2013)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.700950   0.084361   8.309   <2e-16 ***
lnEXP       0.665786   0.009691  68.702   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.59 on 6332 degrees of freedom
Multiple R-squared:  0.4271,	Adjusted R-squared:  0.427
F-statistic:  4720 on 1 and 6332 DF,  p-value: < 2.2e-16
```

$$\widehat{logFDHO}_i = 0.70 + 0.66 logEXP_i$$

Logarithmic Models: Example

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

- The estimate of the elasticity is 0.66. Is that plausible?
- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.
- Food is a normal good, so its elasticity should be positive, but it is a basic necessity.
- Expenditure on it should grow less rapidly than expenditure generally, so its elasticity should be less than 1.
- The intercept has no substantive meaning.

Logarithmic Models: Example

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

- The estimate of the elasticity is 0.66. Is that plausible?
- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.
- Food is a normal good, so its elasticity should be positive, but it is a basic necessity.
- Expenditure on it should grow less rapidly than expenditure generally, so its elasticity should be less than 1.
- The intercept has no substantive meaning.

Logarithmic Models: Example

$$\widehat{logFDHO}_i = 0.70 + 0.66 logEXP_i$$

- The estimate of the elasticity is 0.66. Is that plausible?
- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.
- Food is a normal good, so its elasticity should be positive, but it is a basic necessity.
- Expenditure on it should grow less rapidly than expenditure generally, so its elasticity should be less than 1.
- The intercept has no substantive meaning.

Logarithmic Models: Example

$$\widehat{logFDHO}_i = 0.70 + 0.66 logEXP_i$$

- The estimate of the elasticity is 0.66. Is that plausible?
- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.
- Food is a normal good, so its elasticity should be positive, but it is a basic necessity.
- Expenditure on it should grow less rapidly than expenditure generally, so its elasticity should be less than 1.
- The intercept has no substantive meaning.

Logarithmic Models: Example

$$\widehat{logFDHO}_i = 0.70 + 0.66 logEXP_i$$

- The estimate of the elasticity is 0.66. Is that plausible?
- A 1% increase in total household expenditure leads to a 0.66% increase in food eaten at home expenditure.
- Food is a normal good, so its elasticity should be positive, but it is a basic necessity.
- Expenditure on it should grow less rapidly than expenditure generally, so its elasticity should be less than 1.
- The intercept has no substantive meaning.

Compare Linear and Nonlinear Model

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

- 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 $\beta_1 = e^{0.701} = 2.02$, $\beta_2 = 0.66$

$$FDHO_i = 2.02 EXP_i^{0.66}$$

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

Compare Linear and Nonlinear Model

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

- After log transformation, We estimate the following model:

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

- According to R output,

$$\widehat{logFDHO}_i = 0.70 + 0.66logEXP_i$$

$$log\beta_1 = 0.70, \text{ so } \beta_1 = e^{0.701} = 2.02, \ \beta_2 = 0.66$$

$$FDHO_i = 2.02EXP_i^{0.66}$$

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

Compare Linear and Nonlinear Model

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

- After log transformation, We estimate the following model:

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

- According to R output,

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

$$log\beta_1 = 0.70, \text{ so } \beta_1 = e^{0.701} = 2.02, \ \beta_2 = 0.66$$

$$FDHO_i = 2.02 EXP_i^{0.66}$$

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

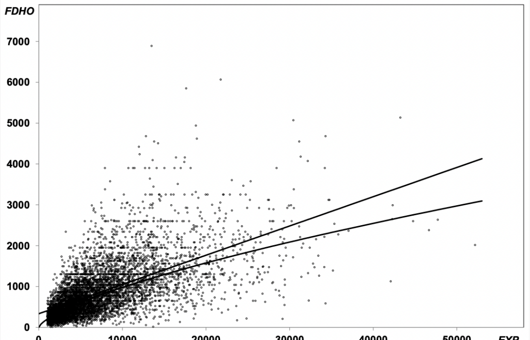
Compare Linear and Logarithmic Model

- Linear model:

$$\widehat{FDHO}_i = 369.44 + 0.063EXP_i$$

- Nonlinear model:

$$FDHO_i = 2.02EXP_i^{0.66}$$



Compare Linear and Logarithmic Model

- Linear model:

$$\widehat{FDHO}_i = 369.44 + 0.063 EXP_i$$

- Nonlinear model:

$$FDHO_i = 2.02 EXP_i^{0.66}$$