ECON 452* -- NOTE 4

Functional Form in the Variables: Linear or Log?

❑ The Natural Logarithmic Transformation

- The ***natural logarithmic transformation*** is **defined only for variables that are *strictly positively valued***.

  $\ln Y_i$ is defined only if $Y_i > 0$ for all i.
  $\ln X_{ij}$ is defined only if $X_{ij} > 0$ for all i.

- **Differentials of $\ln Y_i$ and $\ln X_{ij}$** are the ***relative*, or *proportionate*, changes in Y_i and X_{ij}**.

  $d\ln Y_i = \frac{dY_i}{Y_i}$ = the ***relative (proportionate)*** **change in Y_i** if $Y_i > 0$ for all i.

  $d\ln X_{ij} = \frac{dX_{ij}}{X_{ij}}$ = the ***relative (proportionate)*** **change in X_{ij}** if $X_{ij} > 0$ for all i.

- **Percentage changes in Y_i and X_{ij}** are calculated by multiplying the respective relative changes by 100.

  $100(d\ln Y_i) = 100\left(\frac{dY_i}{Y_i}\right)$ = the percentage change in Y_i if $Y_i > 0$ for all i.

  $100(d\ln X_{ij}) = 100\left(\frac{dX_{ij}}{X_{ij}}\right)$ = the percentage change in X_{ij} if $X_{ij} > 0$ for all i.

- The ***elasticity coefficient*** **of Y with respect to X_j** is defined as follows:

$$\varepsilon_j = \frac{d\ln Y}{d\ln X_j} = \frac{dY/Y}{dX_j/X_j} = \frac{dY}{dX_j}\frac{X_j}{Y}$$ = the ***elasticity*** **of Y wrt to X_j.**

❑ Four Common Functional Form Specifications of Regression Models

***Model 1*: the lin-lin (linear-in-levels) model**

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i \quad \textbf{(1)}$$

- Both the dependent variable Y and the independent variables X_j enter in linear, or linear-in-levels, form.
- The population values of the slope coefficients β_j (j = 1, …, k) depend on the units in which both Y and X_j are measured.

***Model 2*: the log-log (double-log) model**

$$\ln Y_i = \alpha_0 + \alpha_1 \ln X_{i1} + \alpha_2 \ln X_{i2} + \cdots + \alpha_k \ln X_{ik} + u_i \quad \textbf{(2)}$$

- Both the dependent variable Y and the independent variables X_j enter in natural logarithmic, or log, form.
- The population values of the slope coefficients α_j (j = 1, …, k) do not depend on the units in which Y and X_j are measured.

***Model 3*:** **the log-lin (semi-log) model**

$$\ln Y_i = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \cdots + \gamma_k X_{ik} + u_i \qquad \textbf{(3)}$$

- The dependent variable Y enters in log form, but the independent variables X_j enter in linear form.
- The population values of the slope coefficients γ_j (j = 1, …, k) depend on the units in which the X_j are measured, but do not depend on the units in which Y is measured.

***Model 4*:** **the lin-log (inverse semi-log) model**

$$Y_i = \phi_0 + \phi_1 \ln X_{i1} + \phi_2 \ln X_{i2} + \cdots + \phi_k \ln X_{ik} + u_i \qquad \textbf{(4)}$$

- The dependent variable Y enters in linear form, but the independent variables X_j enter in log form .
- The population values of the slope coefficients ϕ_j (j = 1, …, k) do not depend on the units in which the X_j are measured, but do depend on the units in which Y is measured.

❑ **How do these four alternative specifications of the linear regression model differ?**

- These four models are all linear-in-parameters, or linear-in-coefficients. They are all *linear* regression models.
- The only difference among these four models is the interpretation of their respective regression coefficients.

❑ Model 1: the lin-lin model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i \qquad \textbf{(1)}$$

- The ***slope* coefficients β_j (j = 1, …, k)** are interpreted as

$$\beta_j = \frac{\partial Y_i}{\partial X_{ij}} = \frac{\text{change in } Y_i}{\text{change in } X_{ij}}$$ = ***partial slope* of Y with respect to X_j.**

Note: The value of β_j depends on the units in which both Y_i and X_{ij} are measured.

- The ***lin-lin* model** is **linear in the variables Y_i and X_{ij}** (j = 1, ..., k):

partial slope of Y wrt X_j = $\frac{\partial Y_i}{\partial X_{ij}} = \beta_j$ = a constant.

- ***Example 1:***

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$$

$$\beta_1 = \frac{\partial Y_i}{\partial X_{i1}} = \frac{\text{change in } Y_i}{\text{change in } X_{i1}} = \frac{\Delta Y_i}{\Delta X_{i1}}$$ = the ***partial slope* of Y wrt X_1**;

$$\beta_2 = \frac{\partial Y_i}{\partial X_{i2}} = \frac{\text{change in } Y_i}{\text{change in } X_{i2}} = \frac{\Delta Y_i}{\Delta X_{i2}}$$ = the ***partial slope* of Y wrt X_2**.

- ***Example 2:***

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + u_i$$

$$\beta_1 = \frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \frac{\text{change in price}_i}{\text{change in wgt}_i} = \frac{\Delta \text{price}_i}{\Delta \text{wgt}_i}$$

$$\beta_2 = \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \frac{\text{change in price}_i}{\text{change in mpg}_i} = \frac{\Delta \text{price}_i}{\Delta \text{mpg}_i}$$

❑ Model 2: the log-log model

$$\ln Y_i = \alpha_0 + \alpha_1 \ln X_{i1} + \alpha_2 \ln X_{i2} + \cdots + \alpha_k \ln X_{ik} + u_i \qquad \textbf{(2)}$$

- The ***slope*** **coefficients $\boldsymbol{\alpha_j}$ (j = 1, …, k)** are interpreted as

$$\alpha_j = \frac{\partial \ln Y_i}{\partial \ln X_{ij}} = \frac{\partial Y_i / Y_i}{\partial X_{ij} / X_{ij}} = \frac{\partial Y_i}{\partial X_{ij}} \frac{X_{ij}}{Y_i} = \frac{\text{relative change in } Y_i}{\text{relative change in } X_{ij}} = \frac{\%\Delta Y_i}{\%\Delta X_{ij}}$$

= the ***partial elasticity*** **of Y with respect to $\mathbf{X_j}$**.

Note: The value of α_j does not depend on the units in which Y_i and X_{ij} are measured.

- The ***log-log*** **model** is ***nonlinear*** **in the variables $\mathbf{Y_i}$ and $\mathbf{X_{ij}}$** (j = 1, ..., k):

partial slope of Y wrt X_j = $\dfrac{\partial Y_i}{\partial X_{ij}} = \dfrac{\partial \ln Y_i}{\partial \ln X_{ij}} \dfrac{Y_i}{X_{ij}} = \alpha_j \dfrac{Y_i}{X_{ij}}$ = a variable.

- ***Example 1:***

$$\ln Y_i = \alpha_0 + \alpha_1 \ln X_{i1} + \alpha_2 \ln X_{i2} + u_i$$

$\alpha_1 = \dfrac{\partial \ln Y_i}{\partial \ln X_{i1}} = \dfrac{\Delta Y_i / Y_i}{\Delta X_{i1} / X_{i1}} = \dfrac{\text{relative change in } Y_i}{\text{relative change in } X_{i1}}$ = the ***partial elasticity*** **of Y wrt** $\mathbf{X_1}$;

$\alpha_2 = \dfrac{\partial \ln Y_i}{\partial \ln X_{i2}} = \dfrac{\Delta Y_i / Y_i}{\Delta X_{i2} / X_{i2}} = \dfrac{\text{relative change in } Y_i}{\text{relative change in } X_{i2}}$ = the ***partial elasticity*** **of Y wrt** $\mathbf{X_2}$.

- ***Example 2:***

$$\ln(\text{price}_i) = \alpha_0 + \alpha_1 \ln(\text{wgt}_i) + \alpha_2 \ln(\text{mpg}_i) + u_i$$

$$\alpha_1 = \frac{\partial \ln(\text{price}_i)}{\partial \ln(\text{wgt}_i)} = \frac{\text{relative change in price}_i}{\text{relative change in wgt}_i} = \frac{\%\Delta\text{price}_i}{\%\Delta\text{wgt}_i}$$

$$\alpha_2 = \frac{\partial \ln(\text{price}_i)}{\partial \ln(\text{mpg}_i)} = \frac{\text{relative change in price}_i}{\text{relative change in mpg}_i} = \frac{\%\Delta\text{price}_i}{\%\Delta\text{mpg}_i}$$

❑ Model 3: the log-lin model

$$\ln Y_i = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \cdots + \gamma_k X_{ik} + u_i \qquad \textbf{(3)}$$

- The ***slope* coefficients γ_j (j = 1, …, k)** are interpreted as

$$\gamma_j = \frac{\partial \ln Y_i}{\partial X_{ij}} = \frac{\partial Y_i / Y_i}{\partial X_{ij}} = \frac{\partial Y_i}{\partial X_{ij}} \frac{1}{Y_i} = \frac{\text{relative change in } Y_i}{\text{change in } X_{ij}}$$

$$100\gamma_j = \frac{\text{percentage change in } Y_i}{\text{change in } X_{ij}} = \frac{\%\Delta Y_i}{\Delta X_{ij}}$$

= ***partial semi-elasticity* of Y wrt to X_j.**

Note: The value of γ_j does depend on the units in which X_{ij} is measured, but does not depend on the units in which Y_i is measured.

- The ***log-lin* model** is ***nonlinear* in the variable Y_i**:

$$\text{partial slope of Y wrt } X_j = \frac{\partial Y_i}{\partial X_{ij}} = \frac{\partial \ln Y_i}{\partial X_{ij}} Y_i = \gamma_j Y_i = \text{a variable.}$$

- ***Example 1:***

$$\ln Y_i = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + u_i$$

$$\gamma_1 = \frac{\partial \ln Y_i}{\partial X_{i1}} = \frac{\text{relative change in } Y_i}{\text{change in } X_{i1}}; \qquad 100\,\gamma_1 = \frac{\%\Delta Y_i}{\Delta X_{i1}}$$

$$\gamma_2 = \frac{\partial \ln Y_i}{\partial X_{i2}} = \frac{\text{relative change in } Y_i}{\text{change in } X_{i2}}; \qquad 100\,\gamma_2 = \frac{\%\Delta Y_i}{\Delta X_{i2}}$$

- ***Example 2:***

$$\ln(\text{price}_i) = \gamma_0 + \gamma_1 \text{wgt}_i + \gamma_2 \text{mpg}_i + u_i$$

$$100\,\gamma_1 = 100\frac{\partial \ln(\text{price}_i)}{\partial \text{wgt}_i} = \frac{\text{percentage change in price}_i}{\text{change in wgt}_i} = \frac{\%\Delta \text{price}_i}{\Delta \text{wgt}_i}$$

$$100\,\gamma_2 = 100\frac{\partial \ln(\text{price}_i)}{\partial \text{mpg}_i} = \frac{\text{percentage change in price}_i}{\text{change in mpg}_i} = \frac{\%\Delta \text{price}_i}{\Delta \text{mpg}_i}$$

❑ Model 4: the lin-log model

$$Y_i = \phi_0 + \phi_1 \ln X_{i1} + \phi_2 \ln X_{i2} + \cdots + \phi_k \ln X_{ik} + u_i \qquad \textbf{(4)}$$

- The ***slope* coefficients ϕ_j (j = 1, …, k)** are interpreted as

$$\phi_j = \frac{\partial Y_i}{\partial \ln X_{ij}} = \frac{\partial Y_i}{\partial X_{ij} / X_{ij}} = \frac{\partial Y_i}{\partial X_{ij}} X_{ij} = \frac{\text{change in } Y_i}{\text{relative change in } X_{ij}}.$$

$$\frac{\phi_j}{100} = \frac{\text{change in } Y_i}{\text{percentage change in } X_{ij}} = \frac{\Delta Y_i}{\%\Delta X_{ij}}.$$

Note: The value of ϕ_j does depend on the units in which Y_i is measured, but does not depend on the units in which X_{ij} is measured.

- The ***lin-log* model** is ***nonlinear* in the variables X_{ij}** (j = 1, ..., k):

partial slope of Y wrt X_j = $\dfrac{\partial Y_i}{\partial X_{ij}} = \dfrac{\partial Y_i}{\partial \ln X_{ij}} \dfrac{1}{X_{ij}} = \phi_j \dfrac{1}{X_{ij}} = \dfrac{\phi_j}{X_{ij}}$ = a variable.

- ***Example 1:***

$$Y_i = \phi_0 + \phi_1 \ln X_{i1} + \phi_2 \ln X_{i2} + u_i$$

$$\phi_1 = \frac{\partial Y_i}{\partial \ln X_{i1}} = \frac{\text{change in } Y_i}{\text{relative change in } X_{i1}} = \frac{\Delta Y_i}{\Delta X_{i1} / X_{i1}}$$

$$\phi_2 = \frac{\partial Y_i}{\partial \ln X_{i2}} = \frac{\text{change in } Y_i}{\text{relative change in } X_{i2}} = \frac{\Delta Y_i}{\Delta X_{i2} / X_{i2}}$$

- ***Example 2:***

$$\text{price}_i = \phi_0 + \phi_1 \ln(\text{wgt}_i) + \phi_2 \ln(\text{mpg}_i) + u_i$$

$$\phi_1 = \frac{\partial \text{price}_i}{\partial \ln(\text{wgt}_i)} = \frac{\text{change in price}_i}{\text{relative change in wgt}_i}; \qquad \frac{\phi_1}{100} = \frac{\Delta \text{price}_i}{\% \Delta \text{wgt}_i}$$

$$\phi_2 = \frac{\partial \text{price}_i}{\partial \ln(\text{mpg}_i)} = \frac{\text{change in price}_i}{\text{relative change in mpg}_i}; \qquad \frac{\phi_2}{100} = \frac{\Delta \text{price}_i}{\% \Delta \text{mpg}_i}$$

❑ Examples: Four Models of North American Car Prices

Variables:

$price_i$ = the price of the i-th car (in US dollars);
wgt_i = the weight of the i-th car (in pounds);
mpg_i = the miles per gallon (fuel efficiency) for the i-th car (in miles per gallon).

Example 1: a lin-lin model

$$price_i = \beta_0 + \beta_1 wgt_i + \beta_2 mpg_i + u_i \qquad (1)$$

OLS sample regression function:

```
. regress price wgt mpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   14.74
       Model |   186321280     2  93160639.9           Prob > F      =  0.0000
    Residual |   448744116    71  6320339.67           R-squared     =  0.2934
-------------+------------------------------           Adj R-squared =  0.2735
       Total |   635065396    73  8699525.97           Root MSE      =  2514.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   1.746559   .6413538     2.72   0.008      .467736    3.025382
         mpg |  -49.51222   86.15604    -0.57   0.567    -221.3025     122.278
       _cons |   1946.069    3597.05     0.54   0.590    -5226.244    9118.382
------------------------------------------------------------------------------
```

<u>Example 1</u>: a lin-lin model

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + u_i \qquad (1)$$

Interpretation of Slope Coefficient Estimates

$\hat{\beta}_1$ = **1.747** = an estimate of the ***(partial) slope*** of ***price*** with respect to ***wgt***.

- An ***increase*** (decrease) in ***wgt*** of **1 pound** is associated on average with an ***increase*** (decrease) in ***price*** of **1.747 dollars** (per car).

$\hat{\beta}_2$ = **– 49.51** = an estimate of the ***(partial) slope*** of ***price*** with respect to ***mpg***.

- An ***increase*** (decrease) in ***mpg*** (fuel efficiency) of **1 mile per gallon** is associated on average with a ***decrease*** (increase) in ***price*** of **49.51 dollars** (per car).

***Example 2*: a log-log model**

$$\ln(\text{price}_i) = \alpha_0 + \alpha_1 \ln(\text{wgt}_i) + \alpha_2 \ln(\text{mpg}_i) + u_i \qquad (2)$$

OLS sample regression function:

```
. regress lnprice lnwgt lnmpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   15.64
       Model |  3.43231706     2  1.71615853           Prob > F      =  0.0000
    Residual |  7.79121602    71  .109735437           R-squared     =  0.3058
-------------+------------------------------           Adj R-squared =  0.2863
       Total |  11.2235331    73  .153747029           Root MSE      =  .33126

------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnwgt |   .1910324   .2720159     0.70   0.485    -.3513519    .7334167
       lnmpg |  -.6616411   .2784715    -2.38   0.020    -1.216898   -.1063847
       _cons |    9.11759    2.91713     3.13   0.003     3.300998    14.93418
------------------------------------------------------------------------------
```

<u>Example 2</u>: a log-log model

$$\ln(\text{price}_i) = \alpha_0 + \alpha_1 \ln(\text{wgt}_i) + \alpha_2 \ln(\text{mpg}_i) + u_i \quad (2)$$

Interpretation of Slope Coefficient Estimates

$\hat{\alpha}_1 =$ **0.191** = an estimate of the ***partial elasticity*** of ***price*** with respect to ***wgt***.

- A **1 percent *increase*** (decrease) in ***wgt*** is associated on average with a **0.191 percent *increase*** (decrease) in ***price***.
- A **10 percent *increase*** (decrease) in ***wgt*** is associated on average with a 10×0.191 = **1.91 percent *increase*** (decrease) in ***price***.

$\hat{\alpha}_2 =$ **–0.662** = an estimate of the ***partial elasticity*** of ***price*** with respect to ***mpg***.

- A **1 percent *increase*** (decrease) in ***mpg*** is associated on average with a **0.662 percent *decrease*** (increase) in ***price***.
- A **10 percent *increase*** (decrease) in ***mpg*** is associated on average with a 10×0.662 = **6.62 percent *decrease*** (increase) in ***price***.

<u>Example 3</u>: a log-lin model

$$\ln(\text{price}_i) = \gamma_0 + \gamma_1 \text{wgt}_i + \gamma_2 \text{mpg}_i + u_i \qquad (3)$$

OLS sample regression function:

```
. regress lnprice wgt mpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   15.26
       Model |  3.37488699     2  1.68744349           Prob > F      =  0.0000
    Residual |  7.84864609    71  .110544311           R-squared     =  0.3007
-------------+------------------------------           Adj R-squared =  0.2810
       Total |  11.2235331    73  .153747029           Root MSE      =  .33248

------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   .0002087   .0000848     2.46   0.016     .0000396    .0003778
         mpg |  -.0106498   .0113942    -0.93   0.353    -.0333692    .0120696
       _cons |   8.237352   .4757123    17.32   0.000     7.288809    9.185896
------------------------------------------------------------------------------
```

<u>Example 3</u>: a log-lin model

$$\ln(\text{price}_i) = \gamma_0 + \gamma_1 \text{wgt}_i + \gamma_2 \text{mpg}_i + u_i \qquad (3)$$

Interpretation of Slope Coefficient Estimates

$\hat{\gamma}_1$ = **0.0002087**
$100\hat{\gamma}_1$ = **0.02087** = an estimate of the ***partial semi-elasticity*** of ***price*** wrt ***wgt***.

- A **1 pound *increase*** (decrease) in ***wgt*** is associated on average with a **0.02087 percent *increase*** (decrease) in ***price***.
- A **100 pound *increase*** (decrease) in ***wgt*** is associated on average with a 100×0.02087 = **2.087 percent *increase*** (decrease) in ***price***.

$\hat{\gamma}_2$ = **–0.01065**
$100\hat{\gamma}_2$ = **–1.065** = an estimate of the ***partial semi-elasticity*** of ***price*** wrt ***mpg***.

- An ***increase*** (decrease) in ***mpg*** of **1 mile-per-gallon** is associated on average with a **1.065 percent *decrease*** (increase) in ***price***.
- An ***increase*** (decrease) in ***mpg*** of **10 miles-per-gallon** is associated on average with a 10×1.065 = **10.65 percent *decrease*** (increase) in ***price***.

***Example 4*: a lin-log model**

$$\text{price}_i = \phi_0 + \phi_1 \ln(\text{wgt}_i) + \phi_2 \ln(\text{mpg}_i) + u_i \qquad (4)$$

OLS sample regression function:

```
. regress price lnwgt lnmpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   14.36
       Model |   182921988     2  91460994.2           Prob > F      =  0.0000
    Residual |   452143408    71  6368217.01           R-squared     =  0.2880
-------------+------------------------------           Adj R-squared =  0.2680
       Total |   635065396    73  8699525.97           Root MSE      =  2523.5

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnwgt |   1920.301   2072.191     0.93   0.357    -2211.531    6052.132
       lnmpg |  -4332.042   2121.369    -2.04   0.045    -8561.932    -102.152
       _cons |   3946.072   22222.42     0.18   0.860    -40364.17    48256.31
------------------------------------------------------------------------------
```

Example 4: a lin-log model

$$\text{price}_i = \phi_0 + \phi_1 \ln(\text{wgt}_i) + \phi_2 \ln(\text{mpg}_i) + u_i \qquad (4)$$

Interpretation of Slope Coefficient Estimates

$\hat{\phi}_1$ = **1920.3**

$\hat{\phi}_1 / 100$ = **19.20** = an estimate of the ***partial inverse semi-elasticity*** of ***price*** with respect to ***wgt***.

- A **100 percent** ***increase*** (decrease) in ***wgt*** is associated on average with an ***increase*** (decrease) in ***price*** of **1920.30 dollars**.
- A **1 percent** ***increase*** (decrease) in ***wgt*** is associated on average with an ***increase*** (decrease) in ***price*** of **19.20 dollars**.

$\hat{\phi}_2$ = **−4332.0**

$\hat{\phi}_2 / 100$ = **−43.32** = an estimate of the ***partial inverse semi-elasticity*** of ***price*** with respect to ***mpg***.

- A **100 percent** ***increase*** (decrease) in ***mpg*** is associated on average with a ***decrease*** (increase) in ***price*** of **4332 dollars**.
- A **1 percent** ***increase*** (decrease) in ***mpg*** is associated on average with a ***decrease*** (increase) in ***price*** of **43.32 dollars**.