ECON 351* -- NOTE 14

Functional Form in the Variables: Common Specifications

1. Introduction

One of the requirements specified by **Assumption A1** of the CLRM is that the PRE be **linear in coefficients** or **linear in parameters**.

Recall that **linearity-in-coefficients** or **linearity-in-parameters** means that

$$\frac{\partial Y_i}{\partial \beta_j} = f_j(\underline{x}_i)$$, **a function that contains *no unknown parameters***.

But the requirement of linearity-in-parameters does allow the PRE to contain ***known nonlinear* functions** of observable variables Y_i, X_{1i}, X_{2i}, …, X_{ki}.

Two important nonlinear functions in applied regression analysis are (1) the natural logarithmic transformation and (2) the exponential function.

1. The ***natural logarithmic* transformation** is defined only for observable variables that take strictly positive values -- i.e., that are greater than zero for all observations.

Examples: Natural logarithm of $Y_i \equiv \ln Y_i$ if $Y_i > 0$ for all i;
Natural logarithm of $X_{ji} \equiv \ln X_{ji}$ if $X_{ji} > 0$ for all i.

2. ***Known* exponential functions** of observable variables.

Examples: The square (second power) of $X_{ji} \equiv X_{ji}^2$;
The cube (third power) of $X_{ji} \equiv X_{ji}^3$;
The square root of $X_{ji} \equiv X_{ji}^{0.5} = \sqrt{X_{ji}}$.

Note: What is not permitted are ***unknown*** exponential functions of variables, such as X_{ji}^{α} where α is an ***unknown* parameter**.

2. 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.
  $\ln X_{ji}$ is defined only if $X_{ji} > 0$ for all i.

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

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

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

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

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

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

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

$$\varepsilon_j = \frac{d\ln Y_i}{d\ln X_{ji}} = \frac{dY_i / Y_i}{dX_{ji} / X_{ji}} = \frac{dY_i}{dX_{ji}} \frac{X_{ji}}{Y_i} = \text{ the \textbf{elasticity of Y wrt to} } \mathbf{X_j}.$$

3. Three Common Functional Forms for Regression Models

- Two are based on the natural logarithmic transformation of the dependent variable Y_i and/or the explanatory variables X_{ji}.

3.1 The LIN-LIN (Linear) Model

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + u_i \qquad \textbf{(1)}$$

where the ***slope* coefficients β_j (j = 1, 2, …, k)** are interpreted as

$$\beta_j = \frac{\partial Y_i}{\partial X_{ji}} = \frac{\text{change in } Y_i}{\text{change in } X_{ji}} = \text{ partial } \textbf{\textit{slope}} \textbf{ of Y with respect to } \mathbf{X_j}.$$

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

The LIN-LIN model is **linear in the variables Y_i and X_{ji}** (j = 1, 2, ..., k):

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

Example:

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

$$\beta_1 = \frac{\partial Y_i}{\partial X_{1i}} = \frac{\text{change in } Y_i}{\text{change in } X_{1i}} = \text{ the partial } \textbf{\textit{slope}} \textbf{ of Y wrt } \mathbf{X_1};$$

$$\beta_2 = \frac{\partial Y_i}{\partial X_{2i}} = \frac{\text{change in } Y_i}{\text{change in } X_{2i}} = \text{ the partial } \textbf{\textit{slope}} \textbf{ of Y wrt } \mathbf{X_2}.$$

3.2 The LOG-LOG (Double-Log) Model

$$\ln Y_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j \ln X_{ji} + u_i \qquad \textbf{(2)}$$

where the ***slope*** **coefficients α_j (j = 1, …, k)** are interpreted as

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

= the partial elasticity of Y with respect to X_j

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

The LOG-LOG model is **nonlinear in the variables Y_i and X_{ji}** (j = 1, ..., k):

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

Example:

$$\ln Y_i = \alpha_0 + \alpha_1 \ln X_{1i} + \alpha_2 \ln X_{2i} + u_i$$

$\alpha_1 = \dfrac{\partial \ln Y_i}{\partial \ln X_{1i}} = \dfrac{\text{relative change in } Y_i}{\text{relative change in } X_{1i}}$ = the partial ***elasticity*** **of Y wrt X_1**;

$\alpha_2 = \dfrac{\partial \ln Y_i}{\partial \ln X_{2i}} = \dfrac{\text{relative change in } Y_i}{\text{relative change in } X_{2i}}$ = the partial ***elasticity*** **of Y wrt X_2**.

3.3 The LOG-LIN (Semi-Log) Model

$$\ln Y_i = \gamma_0 + \sum_{j=1}^{k} \gamma_j X_{ji} + u_i \qquad \textbf{(3)}$$

where the ***slope* coefficients γ_j (j = 1, …, k)** are interpreted as

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

= the partial semi-elasticity of Y with respect to X_j.

Note: The value of γ_j does depend on the units in which X_{ji} 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**:

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

Example:

$$\ln Y_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + u_i$$

$$\gamma_1 = \frac{\partial \ln Y_i}{\partial X_{1i}} = \frac{\text{relative change in } Y_i}{\text{change in } X_{1i}} = \text{partial } \textbf{\textit{semi-elasticity}} \textbf{ of Y wrt } \mathbf{X_1};$$

$$\gamma_2 = \frac{\partial \ln Y_i}{\partial X_{2i}} = \frac{\text{relative change in } Y_i}{\text{change in } X_{2i}} = \text{partial } \textbf{\textit{semi-elasticity}} \textbf{ of Y wrt } \mathbf{X_2}.$$

$$100\gamma_1 = 100\frac{\partial \ln Y_i}{\partial X_{1i}} = \frac{\text{percentage change in } Y_i}{\text{change in } X_{1i}}$$

$$100\gamma_2 = 100\frac{\partial \ln Y_i}{\partial X_{2i}} = \frac{\text{percentage change in } Y_i}{\text{change in } X_{2i}}$$

4. LIN-LIN, LOG-LOG and LOG-LIN Models: Examples

Example 1. LIN-LIN and LOG-LOG Models of Car Prices

$price_i$ = price of car i, in US dollars

$weight_i$ = weight of car i, in pounds

mpg_i = miles per gallon of car i

LIN-LIN Model of Car Prices

$$price_i = \beta_0 + \beta_1 weight_i + \beta_2 mpg_i + u_i$$

OLS sample regression function:

```
. regress price weight 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]
---------+--------------------------------------------------------------------
  weight |   1.746559   .6413538      2.723   0.008       .4677361    3.025382
     mpg |  -49.51222   86.15604     -0.575   0.567      -221.3025     122.278
   _cons |   1946.069    3597.05      0.541   0.590      -5226.244    9118.382
------------------------------------------------------------------------------
```

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

Meaning: An ***increase*** (decrease) in ***weight*** 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***.

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

LOG-LOG Model of Car Prices

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

OLS sample regression function:

```
. regress lnp lnw lnm

  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

------------------------------------------------------------------------------
     lnp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     lnw |   .1910324   .2720159      0.702   0.485      -.3513519    .7334167
     lnm |  -.6616411   .2784715     -2.376   0.020      -1.216898   -.1063847
   _cons |    9.11759    2.91713      3.126   0.003       3.300998    14.93418
------------------------------------------------------------------------------
```

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

Meaning:

A **1 percent *increase*** (decrease) in ***weight*** is associated on average with a **0.191 percent *increase*** (decrease) in ***price***.

A **10 percent *increase*** (decrease) in ***weight*** is associated on average with a **1.91 percent *increase*** (decrease) in ***price***.

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

Meaning:

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 **6.62 percent *decrease*** (increase) in ***price***.

<u>Example 2</u>. A LOG-LOG Model of Urban Demand for Bus Travel

BUS_i = demand for urban bus travel in city i, in thousands of passenger hours per year

INC_i = average income per capita in city i, in dollars per year

POP_i = population of city i, in thousands of persons

$AREA_i$ = land area of city i, in square miles

Population regression equation (PRE):

$$\ln BUS_i = \alpha_0 + \alpha_1 \ln INC_i + \alpha_2 \ln POP_i + \alpha_3 \ln AREA_i + u_i$$

OLS slope coefficient estimates (standard errors in parentheses):

$\hat{\alpha}_1 =$ **– 4.730** = an estimate of the partial ***elasticity*** of ***BUS*** with respect to ***INC***.
(1.021)

<u>Meaning</u>:

A **1 percent *increase*** (decrease) in ***INC*** is associated on average with a **4.73 percent *decrease*** (increase) in ***BUS***.

$\hat{\alpha}_2 =$ **1.820** = an estimate of the partial ***elasticity*** of ***BUS*** with respect to ***POP***.
(0.236)

<u>Meaning</u>:

A **1 percent *increase*** (decrease) in ***POP*** is associated on average with a **1.82 percent *increase*** (decrease) in ***BUS***.

$\hat{\alpha}_3 =$ **– 0.971** = an estimate of the partial ***elasticity*** of ***BUS*** with respect to ***AREA*.**
(0.207)

<u>Meaning</u>:

A **1 percent *increase*** (decrease) in ***AREA*** is associated on average with a **0.971 percent *decrease*** (increase) in ***BUS***.

Example 3. A LOG-LIN Model of Employees' Wage Rates

W_i = hourly wage rate of employee i, in dollars per hour

ED_i = years of completed schooling of employee i, in years

EXP_i = years of work experience of employee i, in years

Population regression equation (PRE):

$$\ln W_i = \gamma_0 + \gamma_1 ED_i + \gamma_2 EXP_i + u_i$$

OLS sample regression equation (SRE):

```
. regress lnw ed exp

  Source |       SS       df       MS                  Number of obs =     534
---------+------------------------------               F(  2,   531) =   71.17
   Model |  31.3793042     2  15.6896521               Prob > F      =  0.0000
Residual |  117.062591   531  .220456857               R-squared     =  0.2114
---------+------------------------------               Adj R-squared =  0.2084
   Total |  148.441895   533  .278502617               Root MSE      =  .46953

------------------------------------------------------------------------------
     lnw |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ed |   .0964201   .0083113     11.601   0.000        .0800931    .1127472
     exp |   .0117536   .0017542      6.700   0.000        .0083077    .0151996
   _cons |   .5827743   .1254241      4.646   0.000         .336386    .8291627
------------------------------------------------------------------------------
```

$\hat{\gamma}_1 = \mathbf{0.0964}$ $\Rightarrow$ $100\hat{\gamma}_1 = \mathbf{9.64}$

Meaning:

A **1 year *increase*** (decrease) in ***ED*** is associated on average with a 100×0.0964 = **9.64 percent *increase*** (decrease) in ***W***.

$\hat{\gamma}_2 = \mathbf{0.0118}$ $\Rightarrow$ $100\hat{\gamma}_2 = \mathbf{1.18}$

Meaning:

A **1 year *increase*** (decrease) in ***EXP*** is associated on average with a 100×0.0118 = **1.18 percent *increase*** (decrease) in ***W***.