PPOL 503-03, PPOL 503-04, Fall 2016
Course Notes # 2: Nonlinearity and Functional Form

I. Definitions: Linearity and Nonlinearity

Linearity is the assumption that for each independent variable X_i, the amount of change in the mean value of Y associated with a unit increase in X_i, holding all other independent variables constant, is the same regardless of the level of X_i.

In contrast, if for any independent variable X_i in the model, the change in the mean value of Y associated with a unit increase in X_i varies with the value of X_i, we say that X_i is nonlinearly related to the dependent variable.

The relationship between X_i and Y is nonlinear as the slope of the $E(Y)$ curve (the ratio of the change in $E(Y)$ to the change in X_i) resulting from a small increase in X_i varies depending on the level of X_i.

The question is "HOW TO ACCOMMODATE NONLINEARITY?"

If the analyst is convinced that because of theoretical or empirical evidence the relationships among the variables in the model are nonlinear, one needs to find a mathematical specification consistent with the type of nonlinearity expected.

Most nonlinear specifications are nonlinear in terms of both variables and parameters and can only be reasonably estimated using maximum likelihood procedures.

There are a variety of nonlinear specifications that are linear in parameters and for which OLS can be used following an appropriate transformation.

II. Polynomial Models

This nonlinear specification is linear in terms of parameters. It is appropriate for models in which the slope of the relationship between an independent variable X_1 and E(Y) changes sign as the value of X_1 increases

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + u_i$$

The order of this equation is 3; polynomial functions of order 3 are cubic functions.

The graph of the relationship between X_1 and E(Y) expressed by this equation consists of a curve with one or more "bends" - points at which the slope of the curve changes sign. The number of bends = m-1.

Points to Note:

1) The independent variables are each mathematically defined as function of a single conceptual variable. Because the variables are not linearly related, you need not worry about perfect collinearity. The independent variables could be highly correlated so one needs to check for multicollinearity.

2) One can modify the specification to accommodate multiple conceptual variables. Add X_2 and the higher-order terms (squared, cubic, etc.)

3) One needs to be careful in interpreting the estimated coefficients for polynomial models. The typical interpretation of a partial slope coefficient as representing the change in E(Y) associated with a unit increase in an independent variable when all other variables are held constant makes no sense with a polynomial model, as it is impossible for an independent variable to change its value while its higher order powers are held constant.

4) One must interpret regression coefficients for polynomial models by describing the slope of the relationship (and how it changes) over key ranges in the value of the conceptual variable.

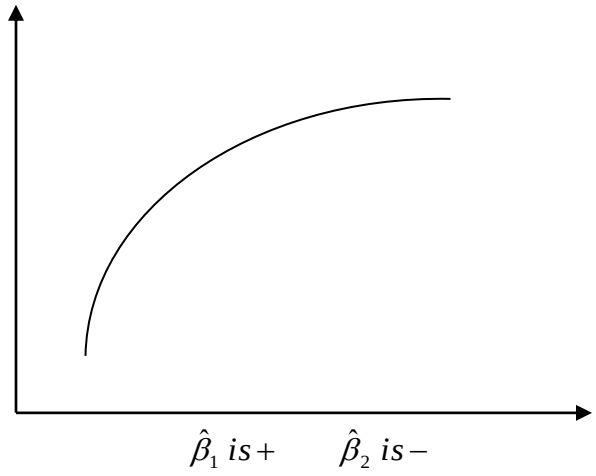
III. QUADRATIC TERMS

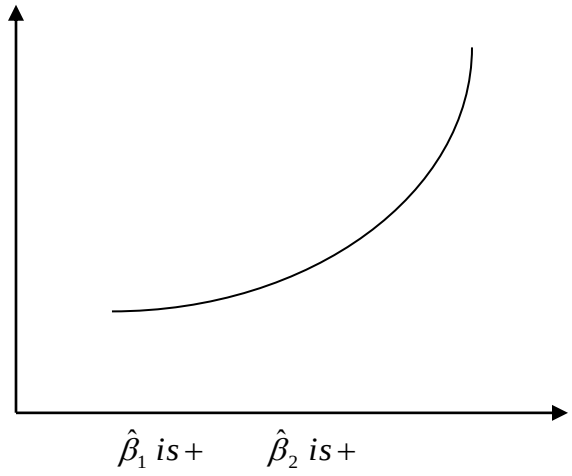
- A *quadratic* indicates a squared form of a variable (X^2). A *cubic* indicates a cubed term (X^3), etc.

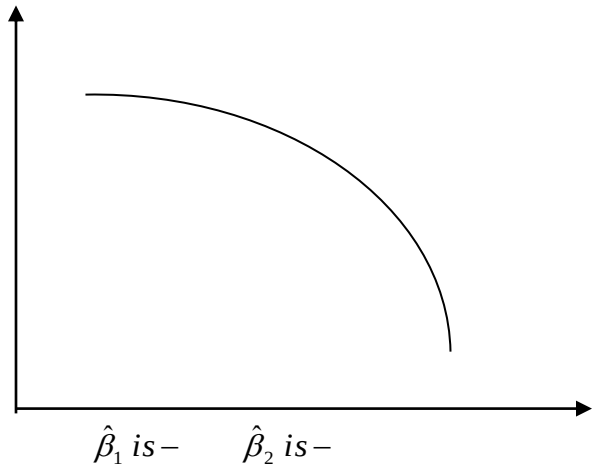
- In models that include quadratics, cubes, etc., it is standard to include all prior transformations also (e.g., if you include X^2, you should also include X; if you include X^3, you should also include X and X^2.

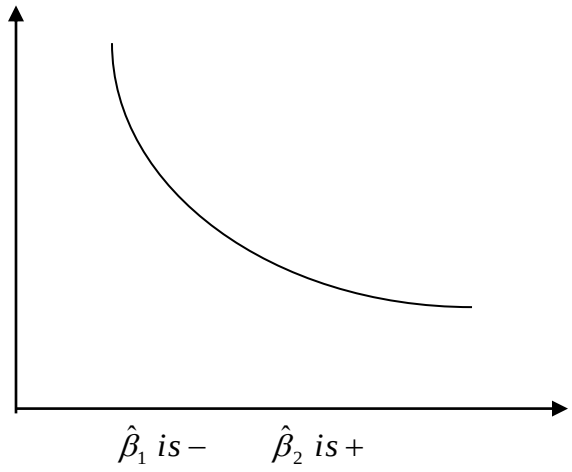
General Model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_1^2$

Pictures:









(data set is CARD.RAW)

```
. sum wage IQ exper expersq

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |      3010    577.2824    262.9583        100       2404
          IQ |      2061    102.4498    15.42376         50        149
       exper |      3010    8.856146    4.141672          0         23
     expersq |      3010    95.57907    84.61831          0        529


. reg wage IQ exper

      Source |       SS       df       MS              Number of obs =    2061
-------------+------------------------------           F(  2,  2058) =   86.71
       Model |  11148498.9     2  5574249.45           Prob > F      =  0.0000
    Residual |   132299572  2058  64285.5065           R-squared     =  0.0777
-------------+------------------------------           Adj R-squared =  0.0768
       Total |   143448071  2060  69634.9861           Root MSE      =  253.55

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   4.604898   .3876702    11.88   0.000     3.844631    5.365164
       exper |   15.29233   1.601728     9.55   0.000     12.15115     18.4335
       _cons |   13.48337   46.64525     0.29   0.773    -77.99344    104.9602
------------------------------------------------------------------------------
```

- Interpretation of the coefficient on EXPER**:** Holding IQ constant, each additional year of experience is associated with a 15.29 cent increase in hourly wages, on average.

- What wage do you predict for a person who has the following number of years of experience? (just use the mean value for IQ in the prediction):

2 years of experience?:	*WAGEHAT* = 13.48 + 4.60(102.45) + 15.29(2) = 515.33 cents/hr
3 years?	*WAGEHAT* = 13.48 + 4.60(102.45) + 15.29(3) = 530.62 cents/hr
20 years?	*WAGEHAT* = 13.48 + 4.60(102.45) + 15.29(20) = 790.55 cents/hr
21 years?	*WAGEHAT* = 13.48 + 4.60(102.45) + 15.29(21) = 805.84 cents/hr

- Now, we'll estimate a model that predicts wage as a function of IQ and experience, where the effects of experience are allowed to vary depending on the level of experience.

  -- One way to do this would be to include an indicator variable for each and every level of experience (except for one baseline category level). This approach requires no assumptions about functional form, but has the disadvantage of requiring many parameters to capture the effect of experience.

  -- An easier and more parsimonious way to model this is to include a squared term of experience, along with the EXPER variable:

```
. reg wage IQ exper expersq

      Source |       SS       df       MS              Number of obs =    2061
-------------+------------------------------           F(  3,  2057) =   64.42
       Model |  12318991.3     3  4106330.42           Prob > F      =  0.0000
    Residual |   131129080  2057  63747.7297           R-squared     =  0.0859
-------------+------------------------------           Adj R-squared =  0.0845
       Total |   143448071  2060  69634.9861           Root MSE      =  252.48

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   4.743112   .3873905    12.24   0.000     3.983394    5.502831
       exper |   41.57254   6.337069     6.56   0.000      29.1448    54.00028
     expersq |   -1.43637   .3352081    -4.29   0.000    -2.093753   -.7789877
       _cons |  -100.1911   53.49143    -1.87   0.061    -205.0941    4.711934
------------------------------------------------------------------------------
```

- In this regression specification, the effect of an extra year of experience depends on the level of experience. So, we can't say what the predicted effect of an extra year of experience is on wages, unless we specify exactly what level of experience we are talking about.
- What wage do we predict now for a person who has

2 years of experience?: *WAGEHAT* = -100.19 + 4.74(102.45) + 41.57(2) –1.44(4) = 462.80 cents/hr
3 years? *WAGEHAT* = -100.19 + 4.74(102.45) + 41.57(3) –1.44(9) = 497.17 cents/hr
20 years? *WAGEHAT* =-100.19 + 4.74(102.45) + 41.57(20) –1.44(400) = 640.82 cents/hr
21 years? *WAGEHAT* = -100.19 + 4.74(102.45) + 41.57(21) –1.44(441) = 623.35 cents/hr

- So, what's the effect of an extra year of experience when that additional year is ***from 2 to 3 years***? We can use two methods. The first one uses the predictions we calculated above:

Method 1: (497.17 – 462.80) = 34.37 more cents/hr

(*Note*: if we would not have rounded the values of the EXPER and EXPERSQ coefficients, we would have calculated exactly a 34.39 cent/hr difference).

Method 2: How could we have estimated this *difference* in predicted wages using just the regression coefficient estimates (i.e., holding constant the values of the other variables in the model)?

$$\Delta WAGEHAT = 41.57254(\Delta EXPER) - (1.43637) * (\Delta EXPERSQ)$$

-- to implement this formula, we need to figure out what the correct values are for $\Delta EXPER$ and for $\Delta EXPERSQ$:

-- The change in experience (i.e. $\Delta EXPER$) in going from 2 to 3 years is just equal to 1 year of experience.

-- However, the change in experience-squared (i.e. $\Delta EXPERSQ$) in going from 2 to 3 years is 9 – 4 = 5 years of experience-squared:

:

$$\Delta WAGEHAT = 41.57254(1) - (1.43637) * (5) = 34.39 \; cents / hr$$

- O.K., let's look at another example: What's the effect of an extra year of experience when that additional year is ***from 20 to 21 years***?

Method 1: (623.35 - 640.82) = -17.47 (Or 17.47 *fewer* cents/hr). If we would not have rounded the values of the EXPER and EXPER squared coefficients, we would have predicted wages of 642.33 and 625.01 respectively, for a difference of 17.32 fewer cents/hr.

Method 2: Or, using the regression coefficients directly, we could have calculated the change in predicted wages as:

-- The change in experience (i.e. $\Delta EXPER$) in going from 20 to 21 years is just equal to 1.

-- However, the change in experience-squared (i.e. $\Delta EXPERSQ$) in going from 20 to 21 years is 441-400 = 41.

-- So: $\Delta WAGEHAT = 41.57254(1) - (1.43637) * (41) = -17.32 \; cents / hr$

Note: We can easily find the "turning point" at which an additional year of experience is predicted to lead to <u>decreased</u> wages by setting the partial derivative with respect to EXPER to zero:

$$\frac{\partial(\hat{Y})}{\partial EXPER} = 41.57 - (2) * (1.44) * (EXPER) = 0$$

$$= -2.88 EXPER = -41.57$$

$$= EXPER = \frac{41.57}{2.88} = 14.43$$

The general formula for computing the turning point is (see Wooldridge page 193):

$$x^* = |\hat{\beta}_1 / (2\ \hat{\beta}_2)|$$

Note that this formula is only relevant when one of the coefficients is positive and the other is negative because the curve does not change directions if both coefficients are the same sign (see pictures on first page of notes).

- So, the return to an additional year of experience becomes negative at 14.43 years of experience. At less than 14.43 years of experience, wages will <u>increase</u> when experience increases. At more than 14.43 years of experiences, wages will <u>decrease</u> when experience increases (try calculating it yourself! at home! and prove to yourself that what the quant swami says is true).

- If we simply ask the question: "conditional on IQ, does experience explain variation in wages," then we have to do a test of <u>joint</u> significance on the coefficients for EXPER and EXPERSQ.

  -- What null is being tested?

  -- How do you interpret these results? Are you surprised? Why or why not?

```
. test exper = expersq = 0

 ( 1)  exper - expersq = 0
 ( 2)  exper = 0

       F(  2,  2057) =   55.14
            Prob > F =    0.0000
```

- **So what's the point?**

-- The estimated wages in using the squared term are lower for each level of experience, compared to the estimates using the coefficient estimates from the regression WITHOUT the squared term.

-- The more important point to see here, however, is that the return (measured in wages/hr) to an extra year of experience depends on the level of experience. In the model with no squared term, an extra year of experience was predicted to increase wages by 15.29 cents/hour, no matter whether that additional year of experience came very early in a career (in this example, from 2 to 3 years), or later in a career (in this example, from 20 to 21 years).

■ By including a squared term for experience along with the level form of experience, we allowed the effect of that extra year of experience to vary depending on the level of experience.

IV. Model Specifications Involving the Logarithmic Function

LOG-LOG: Dependent variable and independent variables are expressed in logs.

$$LN(Y) = \beta_0 + \beta_1 LN(X) + u_i$$

Slope and elasticity are the same.

Example: expenditures on durable goods as a function of total personal consumption expenditures.

$$LN(EXDUR) = -9.69 + 1.91\ LN(PCEX)$$
$$Se = \quad (0.43) \qquad (.051)$$

The elasticity of EXPDUR with respect to PCEX is about 1.91 suggesting that if total personal expenditures increases by 1% then expenditures on durable goods go up by 1.91%.

LOG-LIN (Semi-log): Dependent variable is expressed in logs and independent variable is expressed as actual value.

$$LN(Y) = \beta_0 + \beta_1 (X) + u_i$$

The slope coefficient measures the constant or relative change in Y for a given absolute change in X.

$$\beta_1 = \text{relative change in Y / absolute change in X}$$

If we multiply the relative change in Y by 100, β_1 gives the percentage change or growth rate in Y for an absolute change in X. 100 times β_1 is known as the semi-elasticity of Y with respect to X.

Example: the rate of growth of expenditures on services for the time period 1993 Q1 thru 1998 Q3.

$$\text{LN}(EXS_t) = 7.789 + .00743\,T$$
$$\text{se} = (.00023)\ \ (.00017)$$

over the time period 1993 Q1 thru 1998 Q3, expenditures on services increased at the quarterly rate of .743%.
If the X variables are continuous, then to calculate the elasticity, multiply β_1 by the mean of X and interpret as the elasticity of Y with respect to X.

INTERPRETING INDICATOR VAR COEFFICIENTS IN ln(Y) MODELS

Example 3:

Dependent Variable: lbwght

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.20696	0.20696	5.71	0.0170
Error	1386	50.21338	0.03623		
Corrected Total	1387	50.42034			

Root MSE	0.19034	R-Square	0.0041
Dependent Mean	4.76003	Adj R-Sq	0.0034
Coeff Var	3.99870		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.74730	0.00738	643.18	<.0001
male	1	0.02444	0.01023	2.39	0.0170

- Interpretation of MALE indicator coefficient: Males are predicted to weigh 2.4 percent more at birth, on average, than females ($p = 0.0170$).

- General interpretation of indicator coefficients in predicting ln(Y): [*indicator variable group]* is predicted to [*relevant verb*] $(100 * \hat{\beta})\%$ [*more/less*] than the [*baselinegroup]*

- Note: when the coefficient estimate indicates a relatively "small" percentage change (e.g., as above – say less than 10 percent), then the interpreting the coefficient as above is not far off.

  -- However, when the percentage change indicated is relatively high, then you should calculate an *exact percentage difference* in Y, using the following formula (note: this formula can be used for coefficients of any size – it just produces relatively different estimates when the effect is larger):

  Exact percentage difference in Y for indicator coefficients = $100 * [e^{\hat{\beta}_k} - 1]$

  e.g.: for the male coefficient above: $100 * [e^{0.02444} - 1] = 100 * [1.02474 - 1] = 2.47\%$

If the coefficient on the "male" variable was .34 there would be a greater disparity between the regression coefficient and the percentage change associated with that variable.

$$100 * [e^{.34} - 1] = 100* [1.405 – 1] = 40.5\%$$

In this case male babies are predicted to weigh 40.5% more at birth than female babies.

LIN-LOG: Dependent variable is actual value while the independent variables are expressed in logs.

$$(Y) = \beta_0 + \beta_1 \text{ LN } (X) + u_i$$

The slope measures the absolute change in Y for a percent change in X.

$$\beta_1 = \text{absolute change in Y / change in LN (X)}$$

Since the change in the log of a number is a relative change, it follows that:

$$\beta_1 = \text{absolute change in Y / relative change in (X)}$$

The absolute change in Y is equal to the slope times the relative change in X.

The practical question is when is the LIN-LOG model useful? Engel expenditure model is an interesting application. Engel postulated that total expenditure that is devoted to food tends to increase in arithmetic progression as total expenditure increases in geometric progression.

EXAMPLE: Food expenditure in India as a function of total expenditures.

$$FOODEXP_i = -1283.91 + 257.27 \text{ TOT EXP}$$

$$T\text{-}Stat = \quad (-4.38) \qquad (5.66)$$

The slope coefficient of 257 means that an increase in total expenditures of 1% results in about 2.57 rupees increase in expenditures on food. (note—the slope coefficient is divided by 100).

The elasticity is equal to β_1 / mean (Y).

```
/*******************************************************************************
PPOL 503
Course Notes 2: Quadratics
*******************************************************************************/
cd "C:\...\PPOL509\Stata datasets"

capture: log close
log using "..\do files\notes15.txt", text replace
set more off

clear
use card.dta

sum wage IQ exper expersq

reg wage IQ exper
reg wage IQ exper expersq
test exper = expersq = 0

log close
```