PS2030
Political Research and Analysis
Unit 2: Regression Models: Problems and Extensions
1. Non-Additive and Non-Linear Regression Models

Spring 2025, Weeks 5-6

WW Posvar Hall 3600

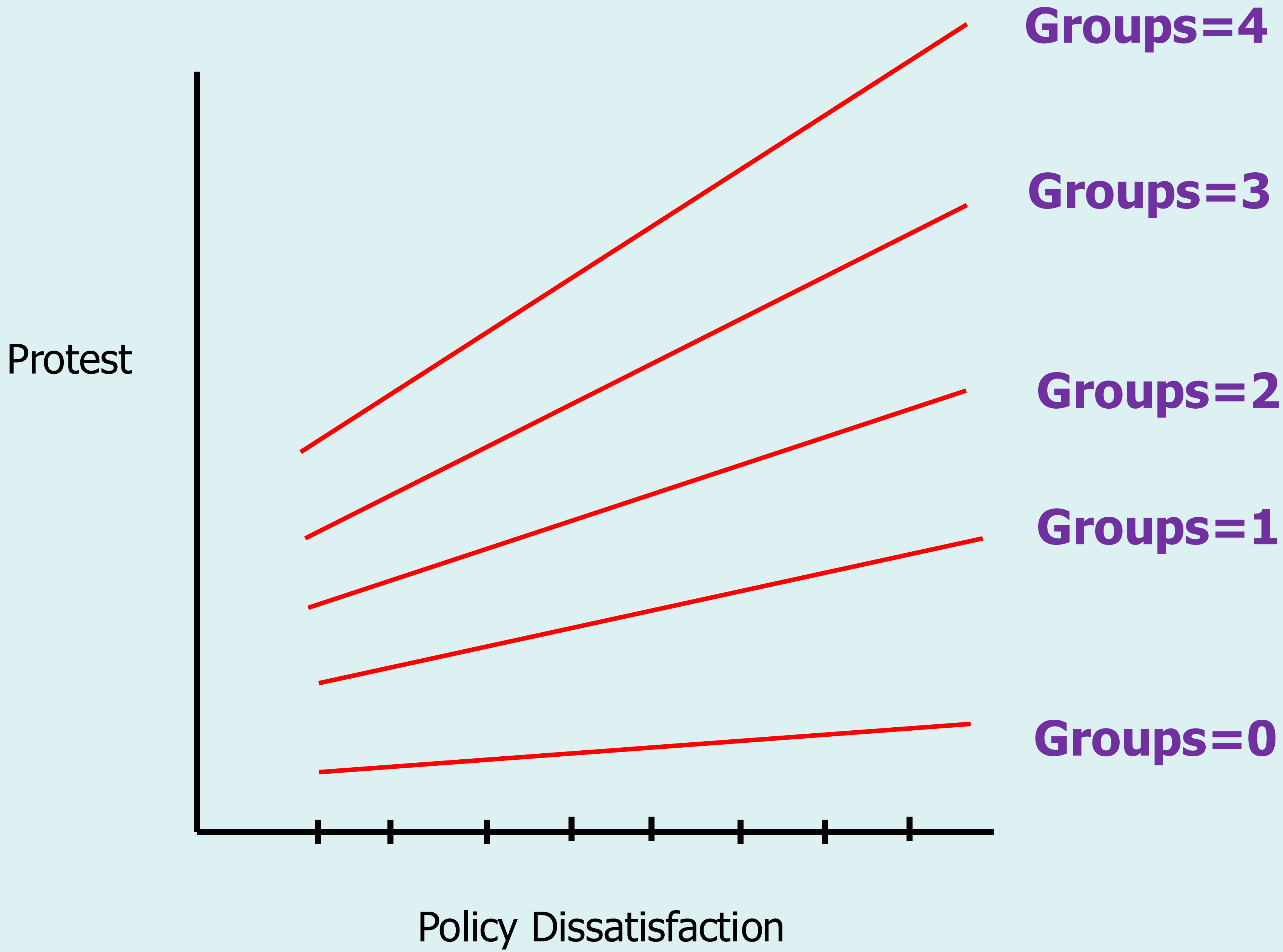
Professor Steven Finkel

Plan for Session

- Extension of "non-additive", interaction or conditional or effects models to include interactions between continuous variables
- Modeling non-linear relationships
- Estimation and interpretation of coefficients in various non-linear regression models

Non-Additive Models

- We can extend the logic of the "slope dummy variable" model to include interactions between two (or more) continuous variables
- In the slope dummy model, the effect of X_1 on Y depended on the group to which an observation belonged (or the category in the dummy variable. This implied two slopes for the effect of X_1 on Y, one for observations in dummy category 0 and the other for observations in dummy category 1
- In the more general case, the effect of variable X_1 on Y depends on the level of a *continuous variable* X_2, and, conversely, the effect of *continuous variable* X_2 on Y depends on the level of variable X_1
- This implies as many different slopes for the effect of X_1 on Y as there are categories in X_2; in theory an infinite number is possible. What is important is that we can use the same general kind of interaction analysis to model exactly how the slope of X_1 on Y may depend on a given unit change in X_2


Groups=4
Groups=3
Protest
Groups=2
Groups=1
Groups=0
Policy Dissatisfaction

- Model is same as slope dummy model with X_2 now being a continuous variable
- Also the case that one almost always has to include both X_2 by itself and the interaction of X_2 and X_1 in the model (as Brambor *et al*. recommend and as common sense dictates)

$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{1i} X_{2i} + e_i$$

where $X_{1i}X_{2i}$ is the product of X_1 and X_2

- So β_3 is the crucial coefficient for the interaction model. It tells you how much the effect of X_1 on Y changes for every unit change in X_2; and it tells you how much the effect of X_2 on Y changes for every unit change in X_1

$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{1i} X_{2i} + e_i$$

$$\textit{where } X_{1i} X_{2i} \text{ is the product of } X_1 \text{ and } X_2$$

- You can see this by rearranging the model in two ways:

$$Y_i = \alpha + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} + \varepsilon_i$$

$$Y_i = \alpha + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} + \varepsilon_i$$

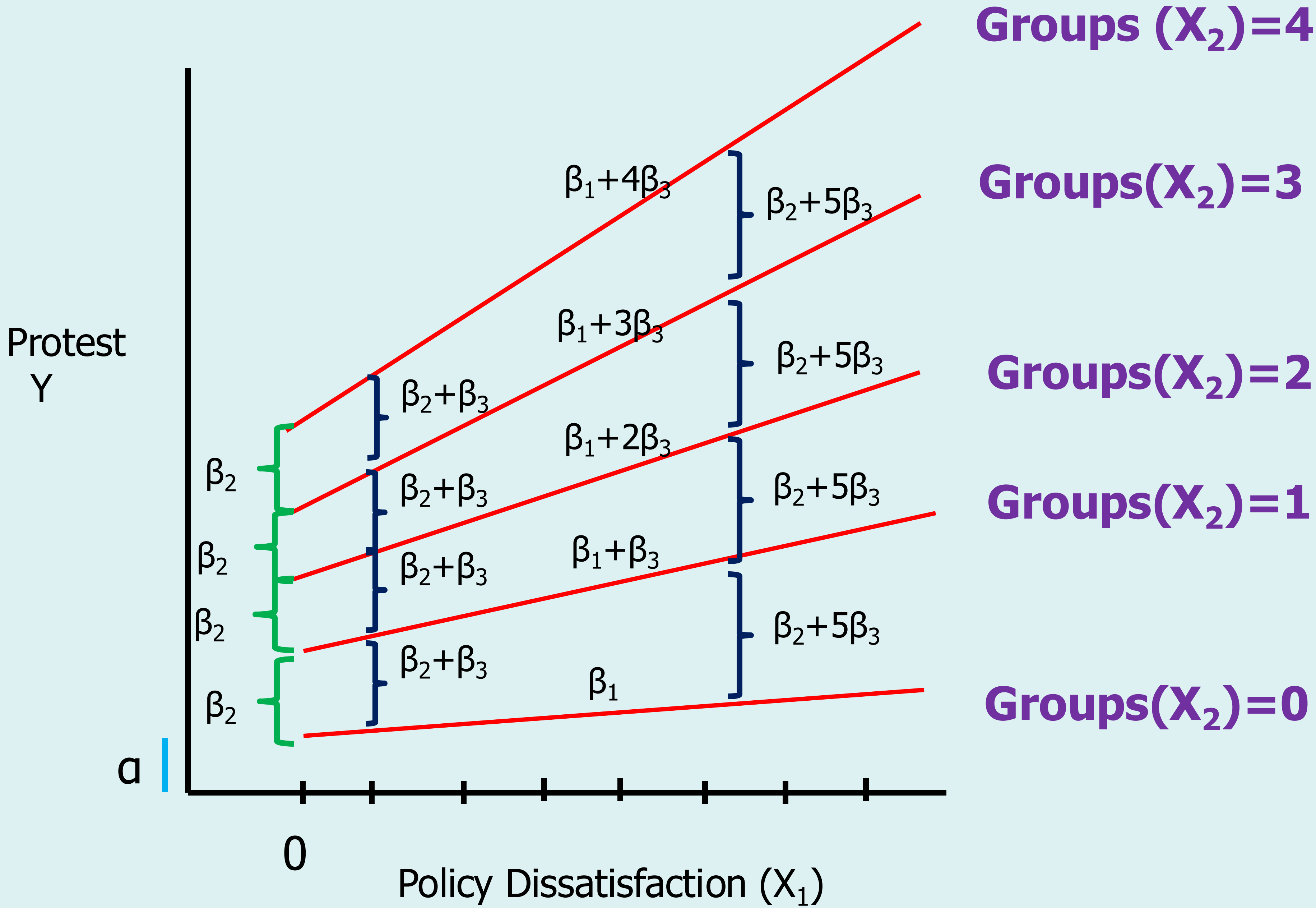
- Implications:
  - The effect of X_2 on Y depends on X_1, increasing by an increment of β_3 for every unit change in X_1
  - The effect of X_1 on Y depends on X_2, increasing by an increment of β_3 for every unit change in X_2
  - **These are equivalent interpretations of this model!!**

- If β_3 is significantly different from 0, then we reject the additive model --- where the effect of each variable is the same at all levels of all other variables --- as holding in the population
- What about β_1 and β_2? They are **not** the same as in an additive model; that is, they are **not** the effect of X_1 and X_2, holding the other constant.
- They are conditional effects of one variable at a single value of the other variable -- when the other variable is 0

$$Y_i = \alpha + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} + \varepsilon_i$$

$$Y_i = \alpha + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} + \varepsilon_i$$

- β_2 is the effect of X_2 on Y when X_1=0
- β_1 is the effect of X_1 on Y when X_2=0
- **This is critical for interpreting interactive models!!!**


Groups $(X_2)=4$
Groups$(X_2)=3$
Groups$(X_2)=2$
Groups$(X_2)=1$
Groups$(X_2)=0$
$\beta_1+4\beta_3$
$\beta_1+3\beta_3$
$\beta_1+2\beta_3$
$\beta_1+\beta_3$
β_1
$\beta_2+5\beta_3$
$\beta_2+5\beta_3$
$\beta_2+5\beta_3$
$\beta_2+5\beta_3$
$\beta_2+\beta_3$
$\beta_2+\beta_3$
$\beta_2+\beta_3$
$\beta_2+\beta_3$
β_2
β_2
β_2
β_2
Protest
Y
α
0
Policy Dissatisfaction (X_1)

- There is nothing necessarily interesting about the slope of X_1 when X_2=0 (and vice versa), so the statistical significance of β_1 and β_2 are not necessarily interesting either
- In fact, if you arbitrarily add a constant to X_1 (or X_2), the zero point will differ anyway, so the significance of the new β_1 and β_2 will potentially differ too
- This means that when thinking about, and presenting, interaction models, you should pick substantively interesting values of X_1 and X_2 and calculate the slopes and their significance at those points
- The slopes at each point will have "conditional standard errors", calculated as:

$$\hat{s}_{b_1|x_{2i}} = \sqrt{\hat{s}^2_{b_1} + \hat{s}^2_{b_3} * X^2_{2i} + 2X_{2i}(Cov(b_1 b_3))}$$

- This is the estimated stand. error of slope of X_1 (β_1) at a specific value of X_2; you can calculate similar conditional standard errors for $\hat{s}_{b_2|x_{1i}}$

Example

```
. gen lrv=vcount*lr

. regress legalcount vcount lr lrv

      Source |       SS       df       MS              Number of obs =     714
-------------+------------------------------           F(  3,   710) =  110.57
       Model |  1562.04199     3  520.680665           Prob > F      =  0.0000
    Residual |  3343.38658   710  4.70899518           R-squared     =  0.3184
-------------+------------------------------           Adj R-squared =  0.3156
       Total |  4905.42857   713  6.87998397           Root MSE      =    2.17

------------------------------------------------------------------------------
  legalcount |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      vcount |   .2566242   .0386094     6.65   0.000     .1808219    .3324265
          lr |   1.089004   .3314736     3.29   0.001     .4382183     1.73979
         lrv |   .0673495   .0446964     1.51   0.132    -.0204033    .1551024
       _cons |   1.373418    .211634     6.49   0.000     .9579143    1.788921
------------------------------------------------------------------------------
```

```
. lincom vcount+lrv*1

 ( 1)  vcount + lrv = 0

------------------------------------------------------------------------------
  legalcount |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3239737   .0492317     6.58   0.000     .2273167    .4206308
------------------------------------------------------------------------------
```

```
. lincom vcount+lrv*5

 ( 1)  vcount + 5*lrv = 0

------------------------------------------------------------------------------
  legalcount |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .5933719   .2147356     2.76   0.006     .1717792    1.014965
------------------------------------------------------------------------------
```

```
. estat vce

Covariance matrix of coefficients of regress model

        e(V) |     vcount          lr         lrv       _cons
-------------+------------------------------------------------
      vcount |  .00149069
          lr |  .00278177   .10987474
         lrv | -.00053235   -.0139903   .00199777
       _cons | -.00743974  -.01982759   .00301955   .04478897
```

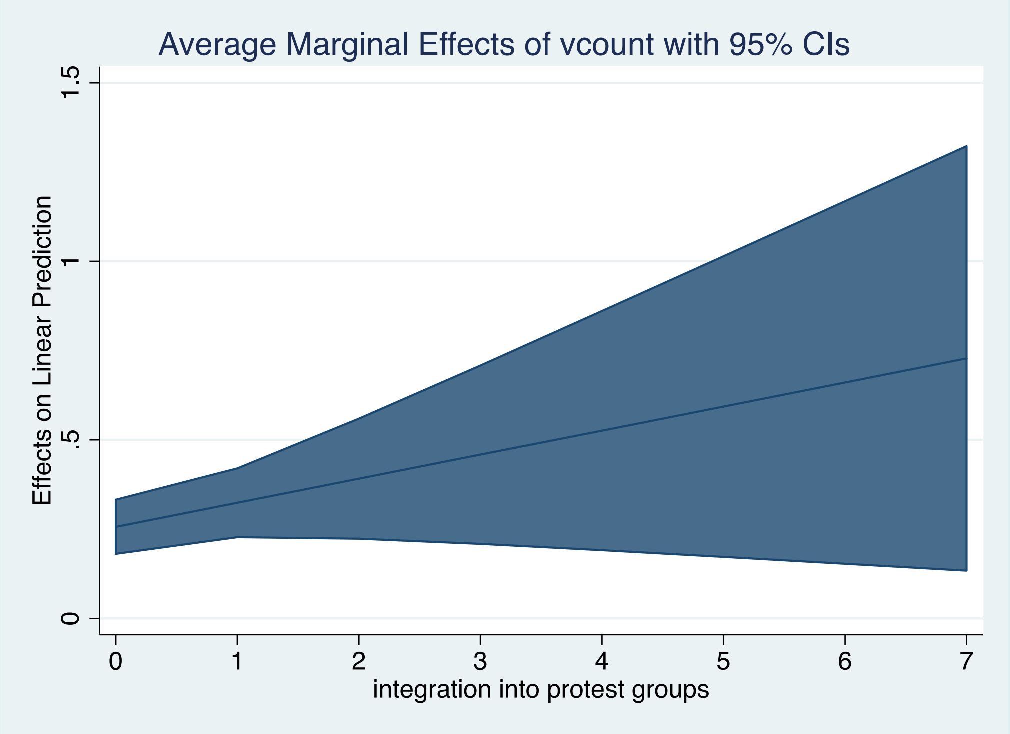
. display (.00149+.00199777+2*-.00053235)^.5 / THE CONDITIONAL STANDARD ERROR WHEN LR=1
.04922469

```
. lincom vcount+lrv*1

 ( 1)  vcount + lrv = 0

------------------------------------------------------------------------------
  legalcount |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3239737   .0492317     6.58   0.000     .2273167    .4206308
------------------------------------------------------------------------------
```

. display (.00149+.00199777*5^2+2*5*-.00053235)^.5 // THE CONDITIONAL STANDARD ERROR WHEN LR=5
.21473414

```
. lincom vcount+lrv*5

 ( 1)  vcount + 5*lrv = 0

------------------------------------------------------------------------------
  legalcount |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .5933719   .2147356     2.76   0.006     .1717792    1.014965
------------------------------------------------------------------------------
```

Graph of the Conditional Relationship with Confidence Intervals (see Stata File for Syntax)



- Look for the point at which the confidence interval overlaps 0; beyond that point the relationship is no longer statistically significant. Here the effect of VCOUNT is significant at all levels of LR, though confidence bands around the estimates at high levels of LR are fairly wide
- Note controversy about whether one can legitimately say that an effect of X_1 is “significant when X_2 is at level 1 but not at level 2” when the interaction effect is not significant

Non-Linear Regression Models

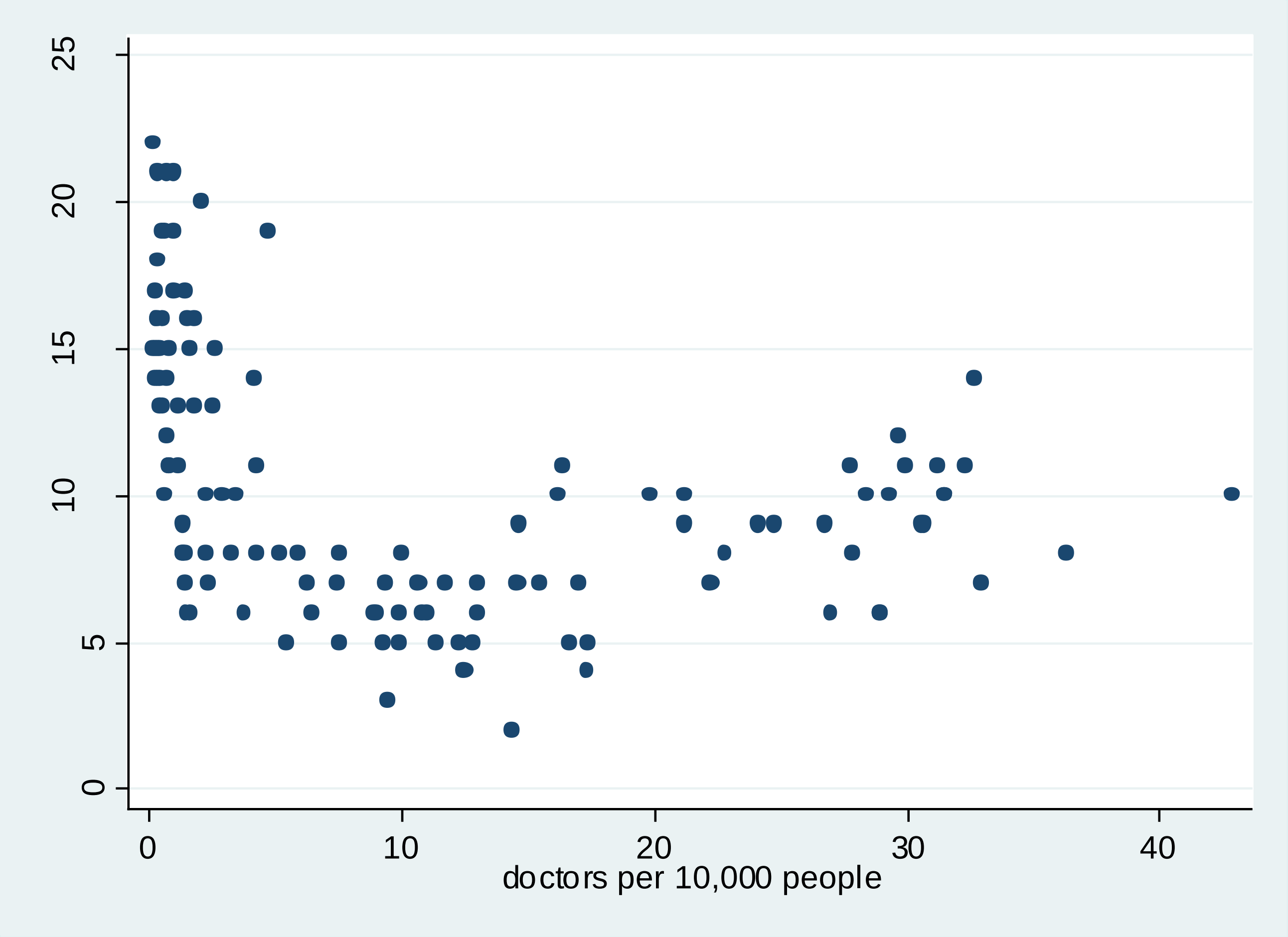
- Linearity Assumption of Ordinary Least Squares: Effect of X on Y is the same, no matter what value of X you choose. Effect of a unit change in X on Y is equal to β, regardless of X changing from 0 to 1, from 10-11, from 50-51, or any other unit change
- This assumption not always accurate, as some relationships between X and Y may be *non-linear*
  - Age (X) and Voter Turnout (Y): As one gets older, the likelihood of voting increases through late middle-age, and then declines as one becomes more elderly. In this case the effect of X on Y is positive up to a certain point, then becomes negative
  - Left-Right ideology (X) and Political Protest (Y): As one moves from extreme left to centrist/moderate, the likelihood of engaging in protest declines, and then increases as one moves from centrist/moderate to extreme right. In this case the effect of X on Y decreases up to a certain point, and then becomes positive

- Country GDP (X) and Average Life Expectancy (Y): As countries become wealthier, this increases life expectancy, but the effect levels off as countries become more and more wealthy due to a biological "ceiling effect" on life expectancy. In this case X affects Y positively, but at a declining rate as X increases until the effect approaches 0 (but never gets there and never turns downward)
- Fertility Rate (X) and Country GDP (Y): As women have more children, wealth in a country declines, but the effect levels off as women have more and more children due to a "floor effect" of GDP. In this case X affects Y negatively, but at an increasing rate as X increases until the effect approaches 0 (but never gets there and never turns upward)

Two Broad Classes of Non-Linear Models

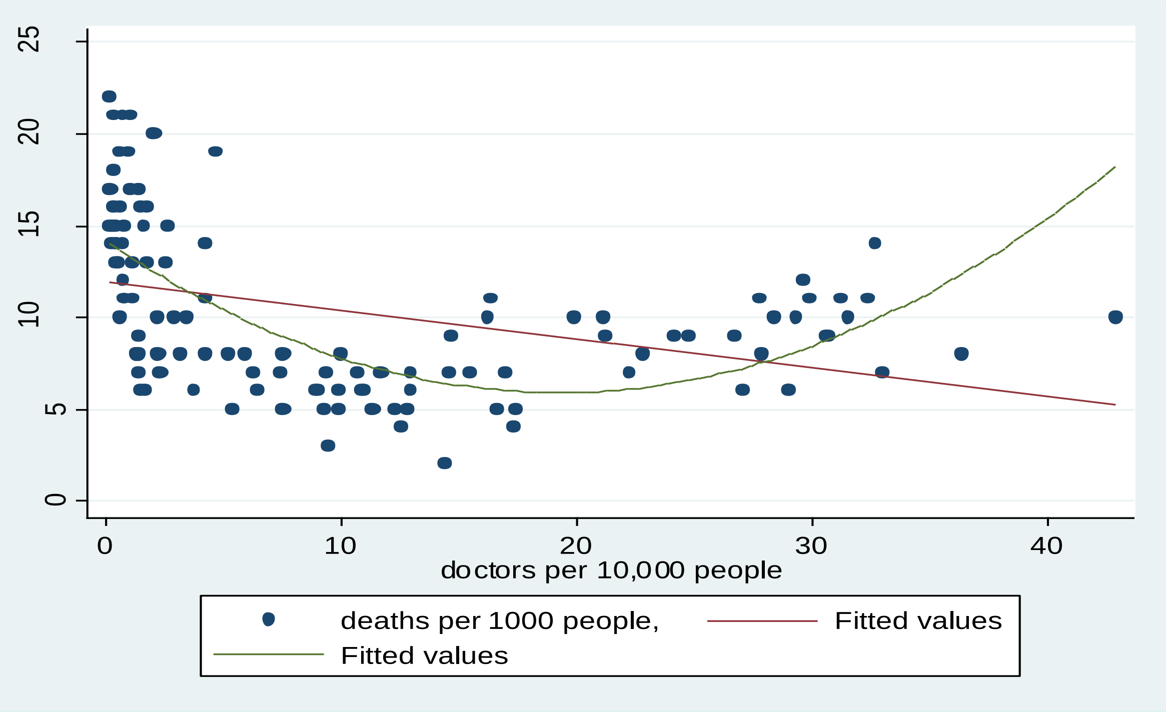
- When the *direction* of the effect of X on Y is thought to change as X increases/decreases, we have a *polynomial* model with additional terms in a new OLS model corresponding to X^2 for the first change of direction, X^3 for the second change of direction, etc.
- When the rate of change of the effect of X on Y is thought to increase/decrease to a *leveling* point without changing direction, we have: either a *logarithmic* model where X and Y are expressed in natural logarithm terms and the model is re-estimated through OLS; or a *hyperbolic* model where a new variable, 1/X, is created and entered into the OLS equation
- In all of these cases we "trick" OLS into estimating a non-linear model by including additional or transformed variables into the usual OLS set-up

Example: Doctors (X) and Country Deathrate (Y)



- Doctors appear to bring down the death rate up to a point, and then increase the death rate! This points to a *polynomial regression model*
- This is probably not a causal effect – probably due to selection effects or reverse causality, whereby underdeveloped countries with very high death rates have few doctors, having some doctors brings down the death rate for many countries, but after that point there are some countries whose death rates are higher and who then attract more doctors
- But let's do the model anyway to see what it looks like and see how to interpret the results of these kinds of models!

Superimpose Linear and Curvilinear Fits on the Graph in STATA



- Curvilinear fits much better (right?)
- Interesting to know: where does the curve change direction, i.e. at what point on X?

Model

$$Y = b_0 + b_1 X_1 + b_2 X_1^2 + e,$$

or

$$Y = b_0 + (b_1 + b_2 X_1) X_1 + e$$

- The effect of X on Y depends on the level of X!! It is a "self-interaction" model!!
- Estimated by generating and including a variable "X_1 squared" in the model

Interpretation of Effects

- If β_1 is positive and β_2 is negative, the relationship between X and Y is positive to a point and then turns downward as the impact of β_2 multiplied by large values of X starts to outweigh the positive impact represented by β_1
- If β_1 is negative and β_2 is positive, the relationship between X and Y is negative to a point and then turns upward
- Using differential calculus, you can solve for the *inflection point* in the relationship (i.e., where the effect is 0) as: $(-\beta_1 / 2\beta_2)$

- If β_2 is significant, along with β_1, then you have a curvilinear relationship that is statistically significant
- Multicollinearity between X_1 and X_1 squared sometimes causes problems, but that's life
- Of course, should control for other variables in the model as well
- If relationship is thought to change direction again, add "X_1-cubed" (X^3) to the equation

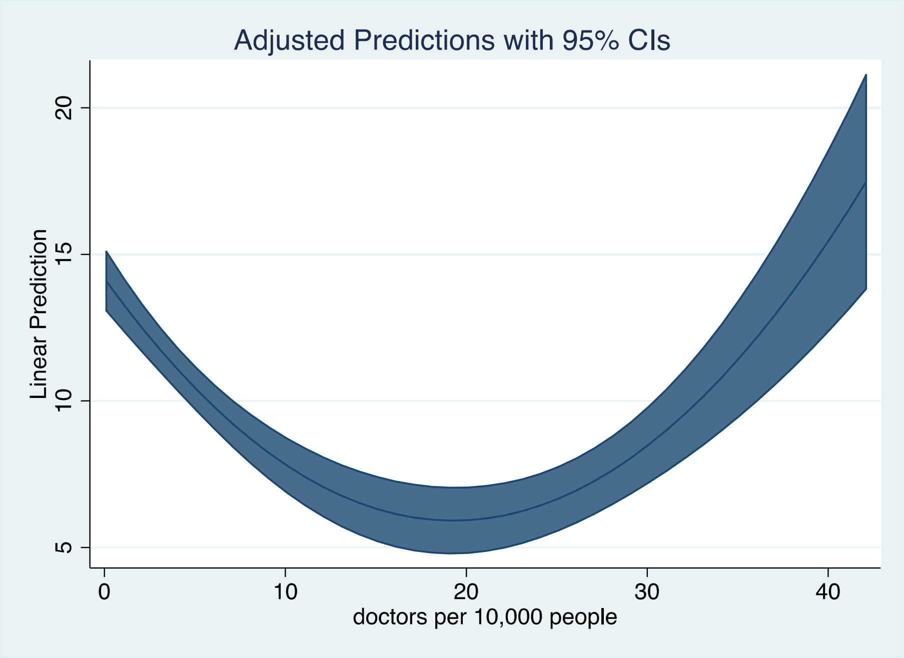
```
. regress deathrat docs docsquared

      Source |       SS       df       MS              Number of obs =     120
-------------+------------------------------           F(  2,   117) =   43.69
       Model |  1044.78622     2  522.39311           Prob > F      =  0.0000
    Residual |  1398.80545   117  11.9556021           R-squared     =  0.4276
-------------+------------------------------           Adj R-squared =  0.4178
       Total |  2443.59167   119  20.5343838           Root MSE      =  3.4577

------------------------------------------------------------------------------
    deathrat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        docs |  -.8564938   .0970364    -8.83   0.000    -1.048669   -.6643183
  docsquared |   .0222062    .002946     7.54   0.000     .0163718    .0280406
       _cons |   14.17583   .5210771    27.20   0.000     13.14387     15.2078
------------------------------------------------------------------------------
```

- β_1 is negative and β_2 is positive. Both are significant. So we have a negative effect of doctors on deathrate, up to a point, after which we have a positive effect.
- The *inflection point* is estimated to be:

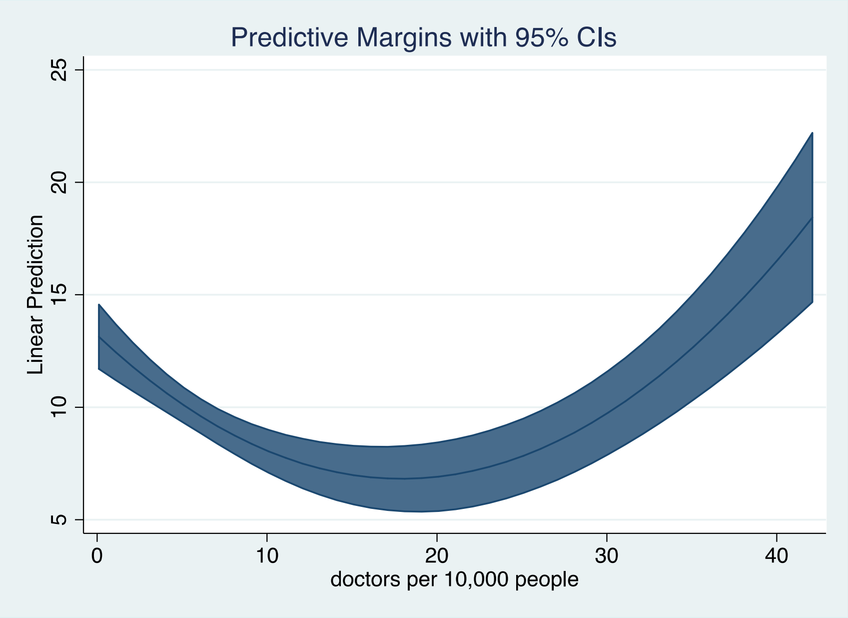
  -(-.856)/(2*.022)= .856/.044 = 19.45
- You can confirm this on the graph on slide 18
- Another example: DEMSAT AND EDUC1 IN SA DATA

Marginsplots in Stata of Curvilinear Relationship

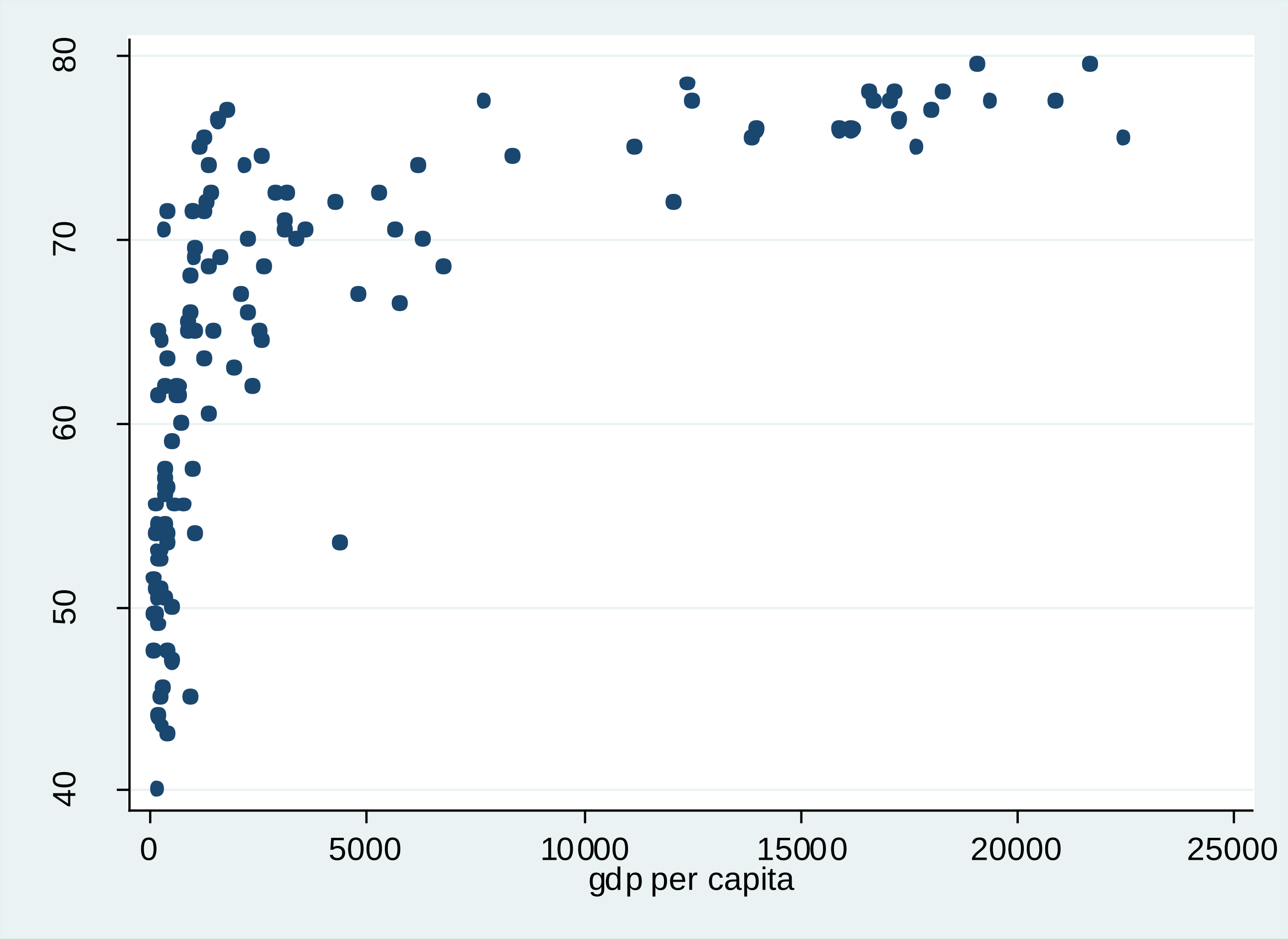
Single Variable with Squared Term



With Additional Covariate held at its Mean



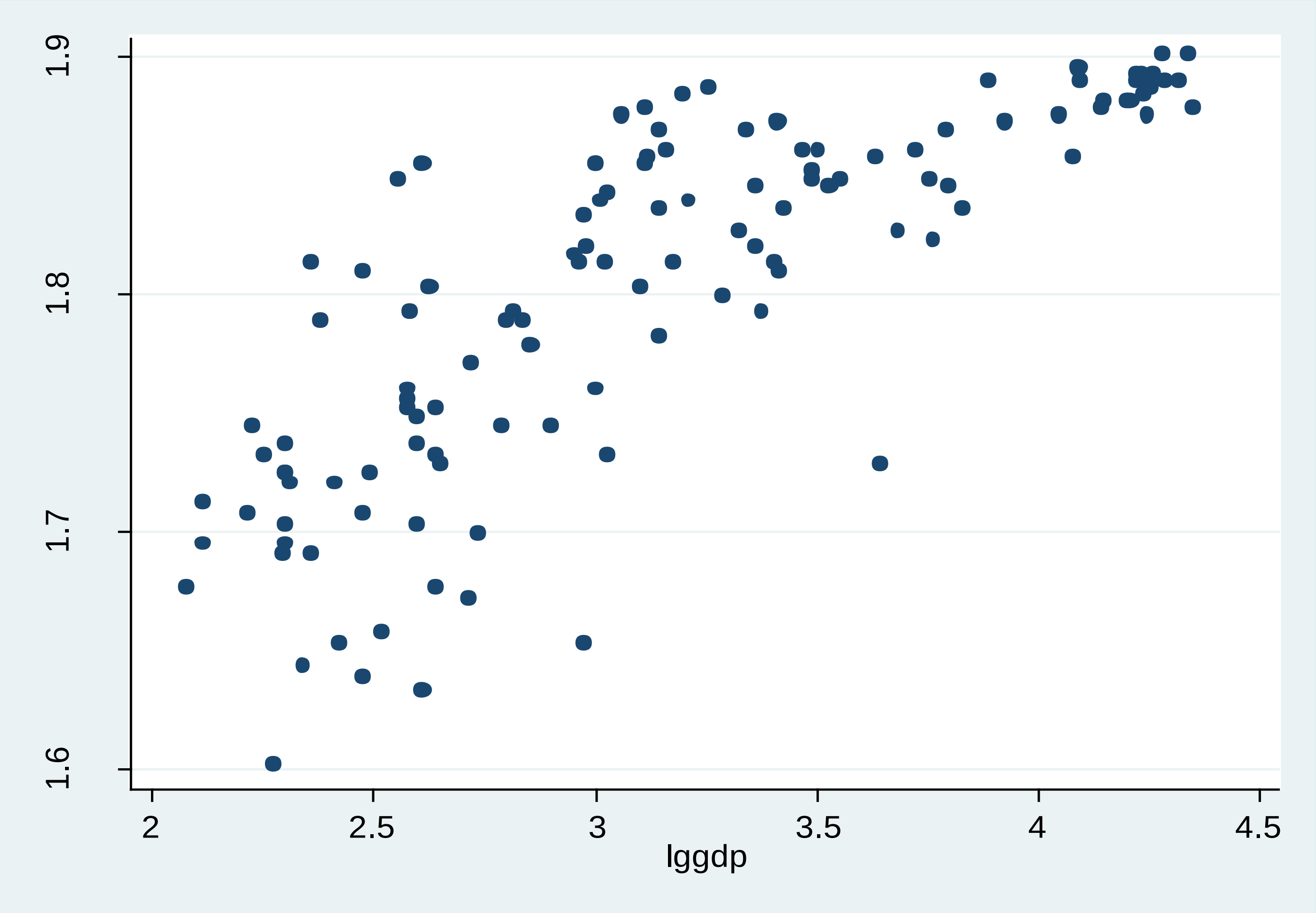
Example: Logarithmic Regression GDP (X) and Life Expectancy (Y)



- We can model this kind of relationship by transforming both X and Y onto a *logarithmic* scale, and re-running the analysis.
- What is the "logarithm" of some variable? Remember logs from high school mathematics?

  - If $10^a = X$, then "a" is the "log of X to the base 10"
    - If X=10, then the log of X to the base 10 is 1, or $\log_{10}(10)=1$
    - If X=100, then the log of X to the base 10 is 2, or $\log_{10}(100)=2$
    - If X=1000, then the log of X to the base 10 is 3, or $\log_{10}(1000)=3$
    - If X=1, then the log of X to the base 10 is 0, or $\log_{10}(1)=0$
    - If X=0, then the log of X is *undefined*!!! There are no logs of 0 or negative numbers. If you have 0s or negative numbers in your data, you need to add a constant to all values (usually "1") so that these cases will not be kicked out in the analysis
  - Logging variables truncates variance (e.g., it put a spread of 10-100-1000 on X onto a 1-2-3 scale instead) and is sometimes used when you have heteroskedasticity in the model as well

So: LGGDP=Log10(GDP) and LGLIFE=Log10(LIFE)



```
. regress lglife lggdp

      Source |       SS       df       MS              Number of obs =     122
-------------+------------------------------           F(  1,   120) =  214.62
       Model |  .444067158     1  .444067158           Prob > F      =  0.0000
    Residual |  .248293825   120  .002069115           R-squared     =  0.6414
-------------+------------------------------           Adj R-squared =  0.6384
       Total |  .692360983   121  .005721992           Root MSE      =  .04549

------------------------------------------------------------------------------
      lglife |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lggdp |   .0917488   .0062628    14.65   0.000     .0793489    .1041487
       _cons |   1.511755   .0201424    75.05   0.000     1.471875    1.551636
------------------------------------------------------------------------------
```

- Coefficient is between 0 and 1: This means that effect starts out positive, decreases in value as X increases, and levels off (decelerating effect)
- Coefficient greater than 1: Effect starts out positive, increases in value as X increases (accelerating effect)
- Coefficient at 1: Effect is just like a linear model

Interpretation of Effects in Log Models

- β_1 is the *elasticity* of Y with respect to X; you can interpret it as the *percentage change in Y for every percentage increase in X*.
- As X and Y get larger, a one percentage point increase in X means more raw units, so the same .09 percentage point increase on Y comes from a larger unit change on X. So the "effect" of X on Y decelerates (so long as β_1 is between 0 and 1).
- So, e.g., going from 10-20 on X is 100% change, and that produces a 9% change in Y from a 10 unit change in X. Going from 100-200 is also a 100% change producing a 9% change in Y from a 100 unit change in X. So much larger unit changes in X produce the same changes in Y as X gets larger, hence the effect "decelerates"

Model

$$Y = \alpha X^{\beta} \varepsilon$$

- which is a fully multiplicative model in the parameters and error term!

- Take the log of both sides and you arrive at:

$$\log Y = \log \alpha + \beta * \log X + \log \varepsilon$$

- Most "log" models do not use "base 10," but rather "base *e*," or the "natural logarithm." The value of *e* is approximately 2.718, so the log(x) to the base *e* is whatever power you have to raise *e* to in order to arrive at the value of X

$$e = 1 + \frac{1}{1} + \frac{1}{1*2} + \frac{1}{1*2*3} + \frac{1}{1*2*3*4} + \frac{1}{1*2*3*4*5} + \ldots.$$

- You can easily add more independent variables to the model, but remember that they enter the formal model multiplicatively so that the transformed model is additive
- You can run "NL" in STATA, which is "non-linear" regression and do the same thing without constructing the new variables. Need to eliminate missing data from consideration first, though.
- (By the way, you can also run NL for the polynomial models earlier!)

- If run log models via the OLS trick, need to remember that R^2 is not directly interpretable since it is "explained variance in the logged units of Y". So generate the "antilog" of the predicted Y in the logged model to get the predicted Y in the real units, and then the square of the correlation between predicted and actual Y gives you "real" R^2

$$e^{\hat{Y}_{\text{logmodel}}} = \hat{Y} \text{ in actual units of Y}$$

then

$$\mathrm{R}^2 = 1 - \frac{\Sigma(Y_i - \hat{Y}_i)^2}{\Sigma(Y_i - \bar{Y}_i)^2} \text{ or}$$

$$\mathrm{R}^2 = r(Y_i, \hat{Y}_i)^2$$

- . nl (life={alpha=0}*gdp^{beta=1})
- (obs = 122)

```
      Source |       SS       df       MS
-------------+------------------------------           Number of obs =        122
       Model |  510339.501     2   255169.75           R-squared     =     0.9911
    Residual |  4601.49945   120  38.3458287           Adj R-squared =     0.9909
-------------+------------------------------           Root MSE      =   6.192401
       Total |      514941   122  4220.82787           Res. dev.     =   789.0952

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |   34.22519   1.495882    22.88   0.000     31.26345    37.18693
       /beta |    .085436    .005598    15.26   0.000     .0743525    .0965196
------------------------------------------------------------------------------
```

- Note near similarity from the manual model with natural logs (in your do file), with e**3.48 power being equal to 32.5 or so

Another Possibility: Hyperbolic Model

- This model is a more extreme version of the log model, in that there is a sharper progression of the effect of X towards leveling off, and Y converging towards some value (the “asymptote”) as X increases more and more
- Look at the graph on slide 24 and you can see that the GDP-LIFE relationship might be better modeled as this kind of process, with the asymptote being something like 75 years of age

Model

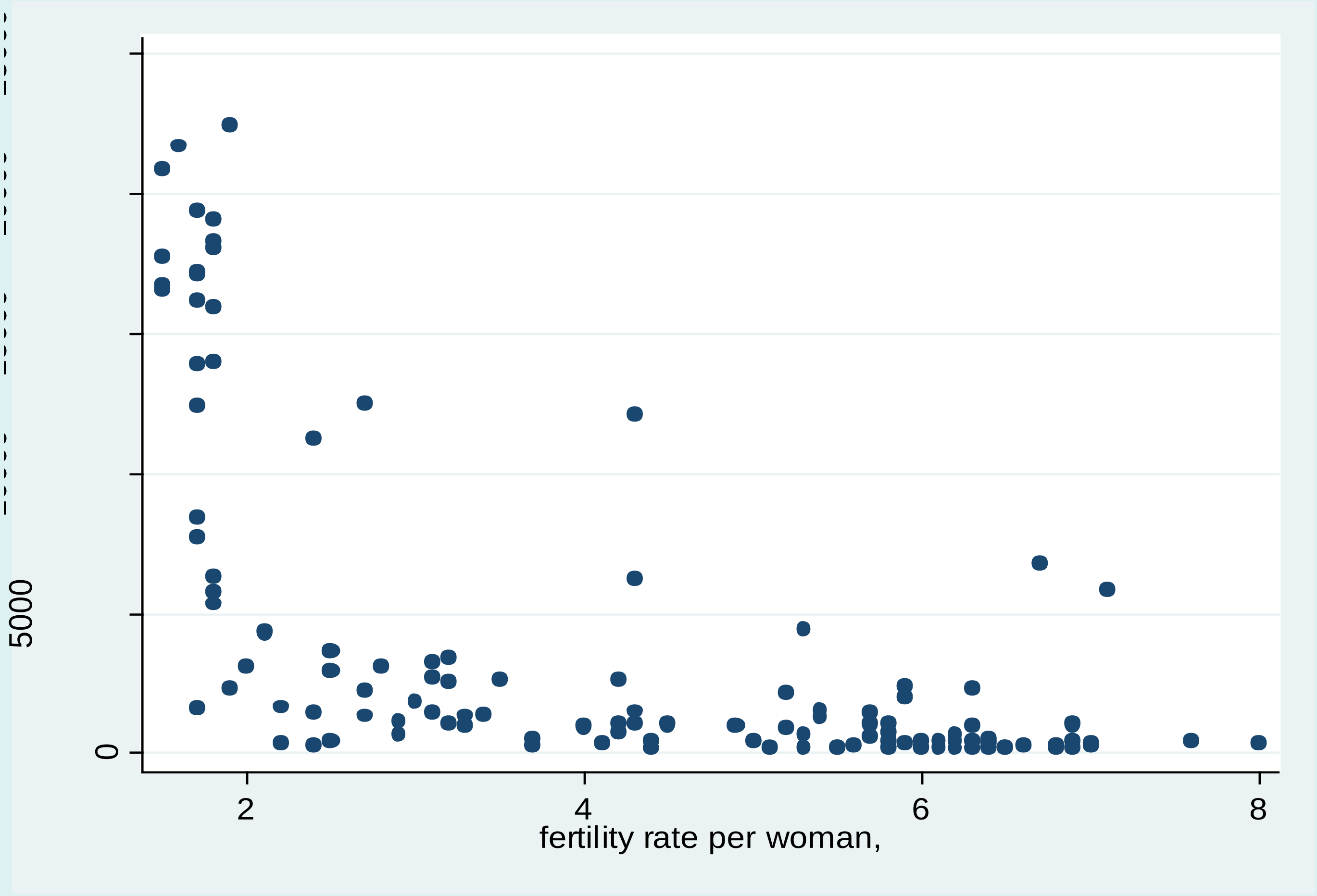
$$Y = \alpha + \beta_1 * \frac{1}{X} + \varepsilon$$

- When X increases to infinity, the whole term goes to 0 and Y converges to the value of the intercept α
- If β is negative, Y approaches α, the asymptote, from below, if β is positive, Y approaches the asymptote from above

Stata Results

```
. nl (life={alpha=70}+{beta=-.5}*1/gdp) if life~=.&gdp~=.
(obs = 122)

Iteration 0:  residual SS =  5966.112
Iteration 1:  residual SS =  5966.112

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =        122
       Model |   7597.5029     1   7597.5029         R-squared     =     0.5601
    Residual |  5966.11185   120  49.7175988         Adj R-squared =     0.5565
-------------+------------------------------         Root MSE      =   7.051071
       Total |  13563.6148   121   112.09599         Res. dev.     =   820.7802

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |   71.06271   .8509818    83.51   0.000     69.37783     72.7476
       /beta |  -4142.371   335.0955   -12.36   0.000    -4805.837   -3478.905
------------------------------------------------------------------------------
  Parameter alpha taken as constant term in model & ANOVA table
```

Example: Fertility Rate (X) and GDP(Y)



Stata Results

```
. nl (gdp={alpha=0}+{beta=1}*1/fertrate) if fertrate~=.&gdp~=.
(obs = 120)

Iteration 0:  residual SS =  1.74e+09
Iteration 1:  residual SS =  1.74e+09
Iteration 2:  residual SS =  1.74e+09

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =       120
       Model |  2.6027e+09     1  2.6027e+09         R-squared     =    0.6000
    Residual |  1.7350e+09   118  14703443.1         Adj R-squared =    0.5966
-------------+------------------------------         Root MSE      =  3834.507
       Total |  4.3377e+09   119  36451632.4         Res. dev.     =  2318.959

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /alpha |  -4427.155   728.7008    -6.08   0.000    -5870.181   -2984.129
       /beta |    28032.7   2106.974    13.30   0.000     23860.31    32205.08
------------------------------------------------------------------------------
  Parameter alpha taken as constant term in model & ANOVA table

.
```