ECON2271

Business Econometrics

(Introductory Econometrics)

Week 3:Topic 2 (ii)

Topic 2(ii): Univariate Regression and OLS

Agenda and learning outcomes

- **Topic 2 (ii): Univariate Regression:** $Y = b_0 + b_1X + e$

  a) Introduce the classic univariate linear regression model

  b) Using univariate simple regression with single dummy to test for differences in group means [Continuous Y, single binary X]

  c) Using OLS to estimating intercept and slope for a linear function [Continuous Y, single continuous X]

  d) The classical assumptions behind OLS:
    - Understand meaning of BLUE estimates;
    - Understand the assumptions about Y and its relationship with X
    - Understand key assumptions about disturbances (*e* or *u*)

  e) Working with nonlinear relationships:
    - Recognise some common nonlinear functional forms, and understand how to specify a function of Y which is linear in the model parameters.
    - Students able to interpret model estimates for simple linear functions, when Y is linear in X, and also some linearized non-linear functions, where Y is not linear in X but is linear in a function of X (i.e. linear in the parameters).
    - Students able to apply techniques required to identify correct functional form and linearized model specification.

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

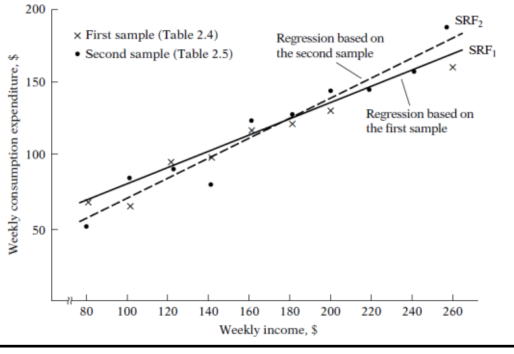
- **OLS and BLUE estimates:**
  - First, a recap: We want to estimate the linear model:

$$Y_i = b_0 + b_1 X_i + e_i$$

  - Y_i: The value of Y for unit (e.g. individual) *i*
  - X_i: The value of X for unit (e.g. individual) *i*
  - e_i: The part of Y_i that is not explained by ($b_0 + b_1X_i$), i.e. the error, disturbance, noise… (=u_i)
    - ➢ We use data for *X* and *Y* to obtain estimates of the parameters b_0 and b_1
    - ➢ The estimator is Ordinary Least Squares (OLS): identifies the straight line of best fit; the set of parameters which minimizes the sum of squared deviations (*e*'s) away from the regression line.
    - ➢ OLS will provide the Best Linear Unbiased and Efficient (BLUE) model estimates, though this is only guaranteed under certain favourable conditions… These conditions are identified by a set of Classical Assumptions of OLS.
  - ❖ Despite the fact that the Classical Assumptions are quite often violated, strictly speaking, OLS tends to be hard to beat. However, that doesn't mean we shouldn't be careful, and any analysis where there is reason to believe that there may be a problem should make attempts to evaluate this and look for solutions, and/or provide some robustness tests.

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

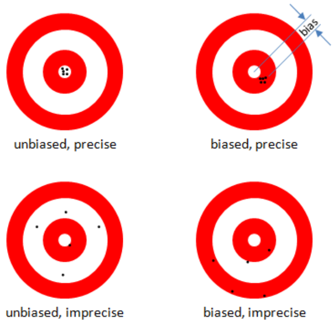
- **A quick reminder about randomness and sampling in the context of regression:**
  - ➢ We estimate the parameters of a model using data drawn from a sample
  - ➢ Hence, we are using the sample data to make inferences (i.e. best possible guesses) about the true value of the model parameters
  - ➢ If we estimate the model based on a different sample, we are unlikely to arrive at exactly the same model estimates, but hopefully they should be close.
  - ➢ Clearly, if we increased our sample, or combined samples, we'd get a better estimate..

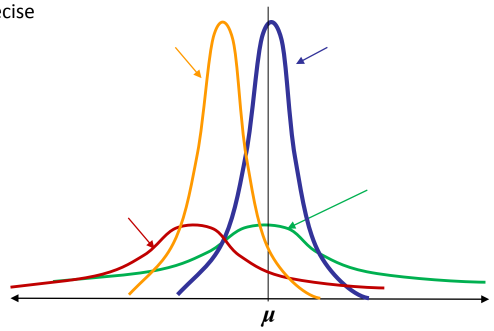


Topic 2: Univariate Regression and OLS

d) The Classical Assumptions of OLS

- **Best Linear Unbiased and Efficient (BLUE) estimates:**
  - ➢ **What do we mean by unbiasedness?**
  - – The estimated sample mean should be close to the population mean
  - – Darts analogy: your aim has good direction
  - ➢ **What do we mean by efficiency? Precision..**
  - – An efficient estimate is precise, i.e. it has a narrow confidence interval
  - – Darts analogy: your throws are precise





Topic 2: Univariate Regression and OLS

d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

❖ **For OLS to yield estimates that are BLUE, the errors must be random and evenly distributed without any systematic patterns:**

- ➢ Errors should be normally distributed around a mean of zero:
- ➢ Errors should have constant variance, which doesn't change with X. That is, errors should be *homoscedastic,* rather than *heteroscedastic.*
- ➢ Errors should be uncorrelated with one another: That is, errors should not be *autocorrelated* or *serially correlated.*
- ➢ Errors should be uncorrelated with X: That is, X needs to be exogenous, rather than endogenous.

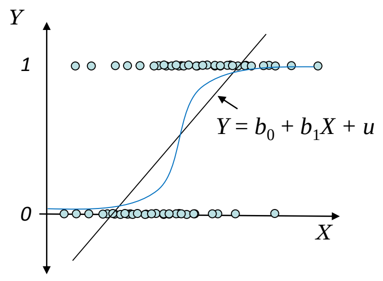
➢ In practice, we avoid problems with the errors by sticking to these rules, as closely as we can:

1. The dependent variable (Y) is (at least approximately) unlimited, continuous and cardinal.
2. We are estimating a model which is linear in its parameters.
3. We are aware of the key symptoms, consequences and solutions for some common problems with errors which tend to arise in specific settings.

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **What if the dependent variable (Y) isn't unlimited, continuous and cardinal?**
  - Consider what happens if Y is binary:
  - E.g: Y = indicator for whether a labour force participant is employed (=1) or unemployed (=0)
    X = Years of schooling



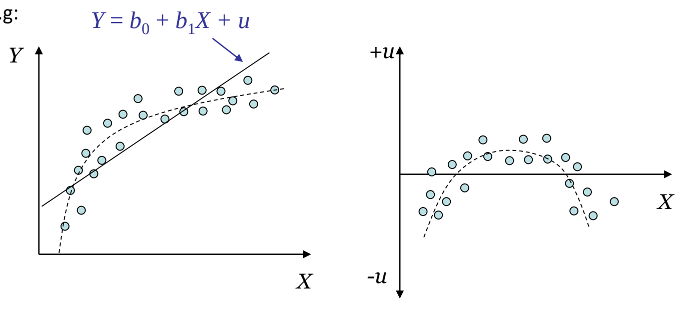
- Clearly, the Y values are lumped in two groups.
- These data are not well spaced around the linear regression line which OLS will produce, and the errors will be huge!
- Consequently, OLS will yield a model with:
  - Large SEs and low R-squared
  - Predicted Y's are continuous will be <0 and >1 for some many values of X.
  - Nonetheless, OLS will probably tell you that there is a positive relationship between X and Y.

  - As you can see, OLS is not well suited to regressions where Y is binary.
  - The solution is to estimate a probability model, which is quite different.
  - We will cover this later in the unit. For now, just be aware of this limitation for our classical linear regression model.
  - Can you think why an ordinal Y would also cause problems? Or censored or truncated Y?

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **What if the relationship between X and Y isn't linear?**
  - If there are nonlinear patterns in the data, but we impose a linear functional form, then our model estimates will clearly not be BLUE…
  - E.g:



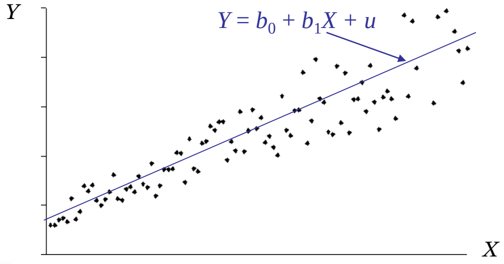
  - In order for OLS to yield BLUE estimates, we require *Y* to be linear in the *model parameters*. Hence, if *Y* does not move linearly with *Y* we may be able to *linearise* the relationship by transforming *X* by some function. In this example it looks like the association between *X* and *Y* might be logarithmic. That means *Y* will be nonlinear in *X*, but it will be linear in ln*X*. More on this later in this lecture…

9

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **"Nonspherical" errors: What if errors are heteroskedastic?**
  - Errors must be homoskedastic, i.e. have constant variance for all X
  - For example, let's say that we are examining how consumption (Y) varies with income (X). We observe a general positive association here, but we might also observe some heteroscedasticity in the data: People with higher incomes may exhibit much more variation in their consumption. This means our ability to predict E(Y|X) will diminish with higher X.
  - There are various tests you can perform to detect heteroscedasticity, but nothing beats having a look: then you can see exactly where the variances are blowing out too, which is useful.

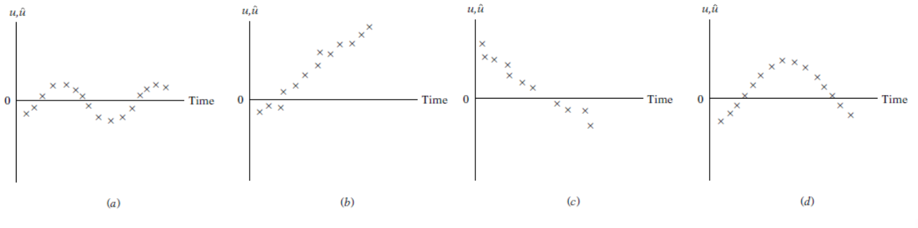


10

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **"Nonspherical errors: What if the errors are correlated with one another?**
  - Means there are patterns in the error terms: cov (u_i, u_j) $\neq$ 0
  - Often referred to as autocorrelation or serial correlation.
  - These problems arise most often in time-series data, which we haven't covered yet, but:
  - In a cross-sectional setting we might find errors "spilling over" to other affected units: if I'm less happy than I should be, given my other observable characteristics, chances are the rest of my household is too. So our errors are correlated.



11

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Consequences and solutions for nonspherical errors:**
  - Happily, nonspherical errors do NOT cause OLS estimates to be biased.
  - However, variances are no longer reliable, which means hypothesis testing won't be reliable either.
    - But, happily, there are ways to address that problem, and modern statistical packages make this easy: you can usually select to estimate "robust standard errors", and all is well.
  - Further, OLS estimates won't necessarily be the most efficient, producing the parameters with the lowest variance.
    - But, happily, there are ways to address this problem too: There is another estimator which will be BLUE, called Generalised Least Squares (GLS) estimator. Instead of minimising the sum of squared residuals, GLS minimises an appropriately *weighted* sum of squared residuals. (And you can select the appropriate weighting method to suit your specific problem.)

12

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Consequences and solutions for nonspherical errors: A caveat!**
  - Later in the unit we will consider multivariate regression models, where there are more than one explanatory (right-hand-side) variable. Then we will be able to specify models where Y can be explained by a combination of variables. This is very exciting, but with all fun things comes responsibility: the problem of correct model specification becomes more complex. Not only do we have to worry about linearity issues, we also have to worry about selecting the correct combination of variables.
  - So, before we try to fix a problem with errors, Kennedy (in his text, chapter 5) reminds us to stop and think:

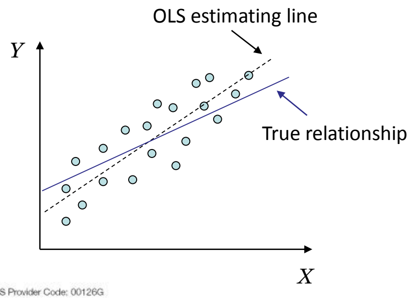
    "Funny-looking errors should at first be interpreted as signalling a specification error, not a nonspherical error".

  - As always is the case: the best way to fix a problem is to find the root cause of it, and fix it there. Don't just fix the symptoms, unless it's your last option.
  - Just because your software provides lots of fancy options and correction methods in your estimation procedure doesn't mean you should take out the full artillery. You should make sure you understand what it does and why, and you should remember that by trying to fix one problem you can create another one which is even bigger…

13

Topic 2: Univariate Regression and OLS
d) The Classical Assumptions of OLS

- **Endogeneity: What if the errors are correlated with X?**
  - The consequences of endogeneity is biased estimates, and this is much harder to circumvent and fix than problems associated with "nonspherical" errors.
  - It is, essentially, a mostly invisible problem which does not reveal itself by funny-looking errors or any other easily observable feature. Instead, we have to determine whether it is likely to be a problem, in any given setting, using logic, theory, intuition and evidence. If we can't make a good case for why we don't have a problem, then we have to assume that we have one, and try to deal with it as best we can.



There are 3 main root-causes of endogeneity:

1. Omitted variables
2. Simultaneity and reverse causality
3. Measurement error

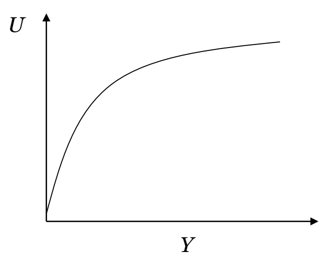
The problem of endogenous regressors is complex and difficult to explain in simple terms, so we will return to this later.

14

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **What do we do if we are working with nonlinear functions?**
  - Many relationships we work with in Economics are nonlinear…
  - E.g. utility functions
    - U = some measure of utility
    - Y = some measure of income
  - Theory predicts:

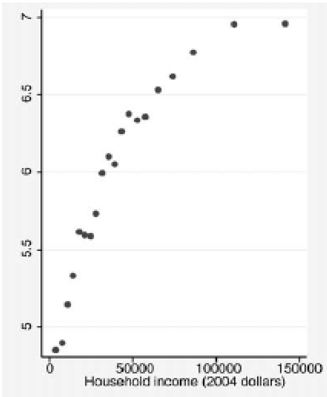


CRICOS Provider Code: 00126G

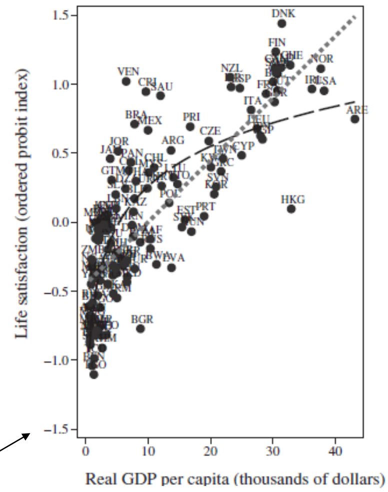
15

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

– Empirical evidence confirms:



R. Layard et al. / Journal of Public Economics 92 (2008) 1846–1857

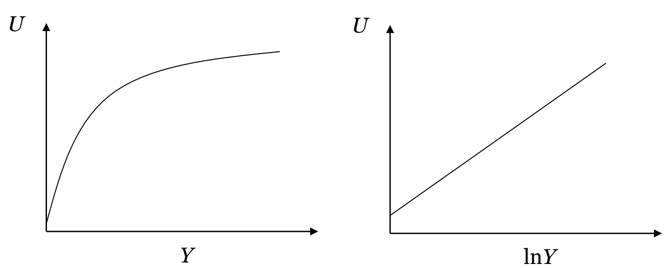


Stevenson, B. and & J. Wolfers (2008) Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. *Brookings Papers on Economic Activity,* Spring 2008

16

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

- Many of the relationships we study in economics are, in fact, logarithmic.
- In this case: $U_i = b_0 + b_1(\ln Y_i) + e_i$
  - ➢ This is called the lin-log specification, because it proposes that U is linear in the log of Y.
    - ➢ Other possible relationships imply log-lin and log-log specifications.
    - ➢ What happens when we map U_i on $\ln Y_i$ instead of on Y_i ?
    - ➢ Theory predicts:



17

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

– Empirical evidence confirms:

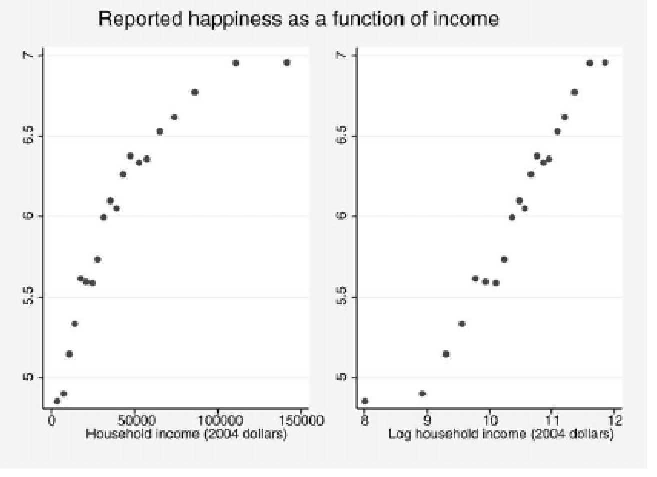
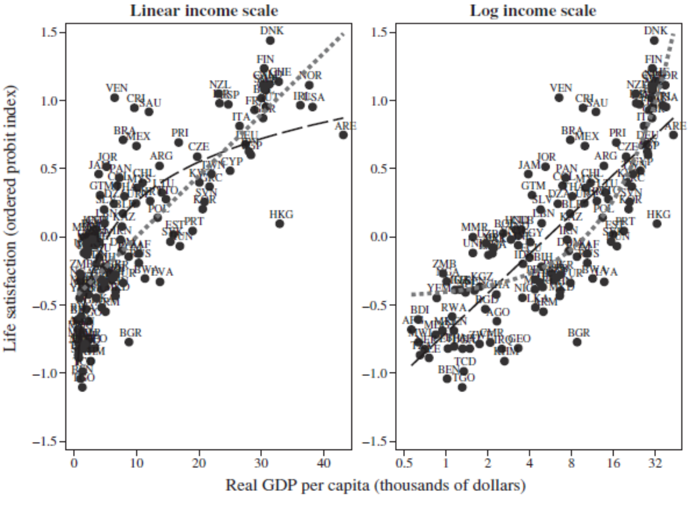


Fig. 2. The simple cross-sectional relationship between reported happiness and reported income and between reported happiness and log reported income in the U.S. General Social Survey.

18

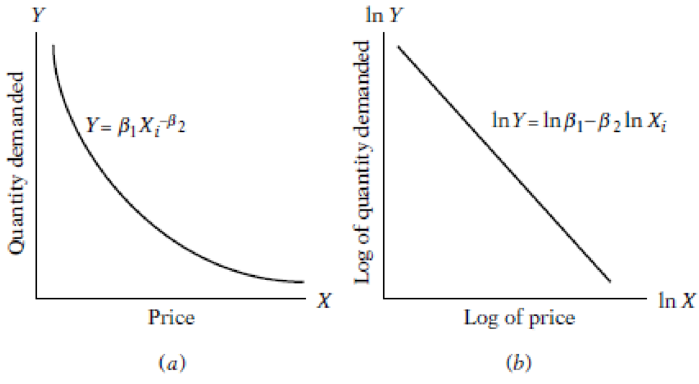
Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

Figure 7. Assessing the Functional Form of the Life Satisfaction–GDP Gradient: Gallup World Poll[a]



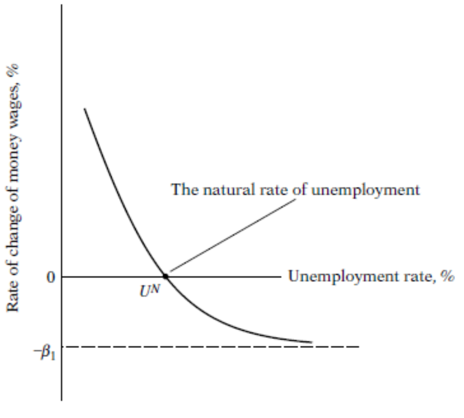
Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

- **Other types of nonlinear functional forms**
  - We find other types of nonlinear relationships too...
    - E.g. The log-log functional form:



Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

- **Other types of nonlinear functional forms**
  - E.g. the reciprocal (hyperbolic) functional form:

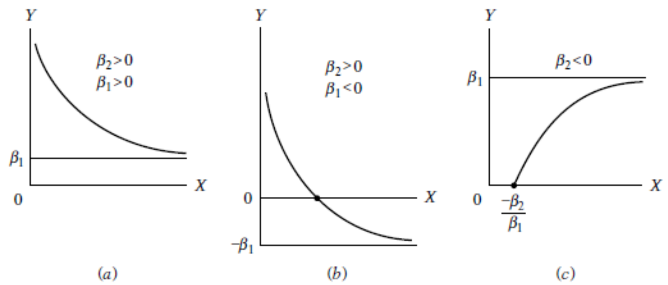


The Phillips curve.

21

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

- **Other types of nonlinear functional forms**
  - Reciprocal (hyperbolic) functional forms:



The reciprocal model: $Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$.

22

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

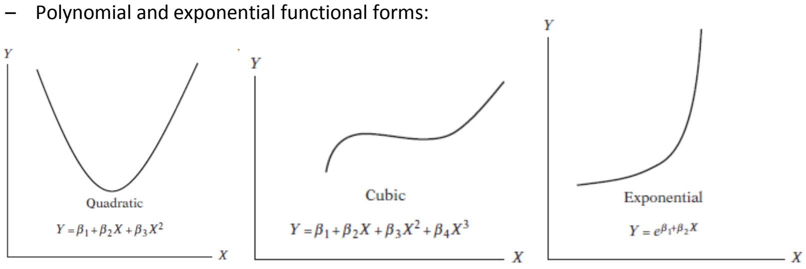
- **Other types of nonlinear functional forms**
  - An overview of variants of logarithmic and reciprocal (hyperbolic) functional forms:

Model	Equation	Slope $\left(=\frac{dY}{dX}\right)$	Elasticity $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$
Linear	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
Log–linear	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log–lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 (Y)$	$\beta_2 (X)^*$
Lin–log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
Reciprocal	$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
Log reciprocal	$\ln Y = \beta_1 - \beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

Note: * indicates that the elasticity is variable, depending on the value taken by X or Y or both. When no X and Y values are specified, in practice, very often these elasticities are measured at the mean values of these variables, namely, $\bar{X}$ and $\bar{Y}$.

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

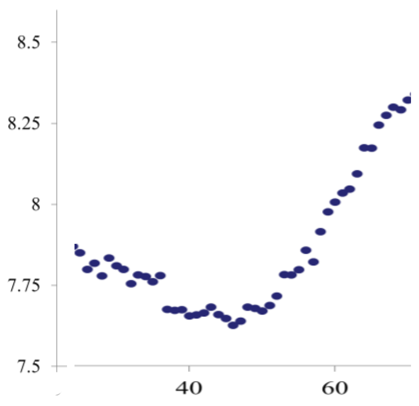
- **Other types of nonlinear functional forms**
  - Polynomial and exponential functional forms:



$\frac{dY}{dX} = \beta_2 + 2\beta_3 X;$ $\frac{dY}{dX} = \beta_2 + 2\beta_3 X + 3\beta_3 X^2$ $\frac{dY}{dX} = \beta_2 e^{\beta_1 + \beta_2 X}$

Topic 2: Univariate Regression and OLS
e) Working with nonlinear relationships

- **Other types of nonlinear functional forms**
  - E.g. What is the relationship between age and life satisfaction?
    - Sample 1: People of "working age"

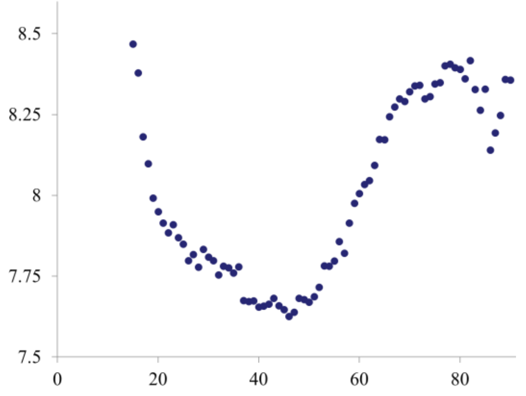


Topic 2: Univariate Regression and OLS

e) Working with nonlinear relationships

The University of Western Australia
Achieve International Excellence

- **Other types of nonlinear functional forms**
  - E.g. What is the relationship between age and life satisfaction?
    - Sample 2: All ages



Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

- **The general approach to working with nonlinear patterns**
  - ➢ The important condition for OLS is that Y is linear *in the model parameters*:
    - ➢ E.g. for the lin-log model, Y is NOT linear in X, but it is linear in lnX
    - ➢ E.g. for the quadratic model, Y is NOT linear in X, but it is linear in the X and X^2
  - ➢ As long as Y is linear in the parameters, we can estimate the model using OLS
  - ➢ Therefore it is important to establish the correct functional form
  - ➢ How do you know?
    - ➢ Look to the theory: e.g. the Phillips Curve; the Utility Curve
    - ➢ Look at the data: Can you see any nonlinear patterns?
    - ➢ Check for departures from linearity by adding a squared term (X^2) to a standard linear model.

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

- **The general approach to working with nonlinear patterns**
  - Once you have discovered a departure from linearity in your data, you need to figure out what type of nonlinearity
    - Just because the coefficient for X^2 is significant doesn't mean the quadratic specification is best..
    - A quadratic specification is great for picking up different types of nonlinearity, which can be:
      - Positive monotonic, i.e. increasing for all feasible values of X, but not necessarily at the same rate;
      - Negative monotonic, i.e. decreasing for all feasible values of X, but not necessarily at the same rate;
      - Concave up, or U-shaped, i.e. it has a minimum turning-point; or
      - Concave down, or inverted-U-shaped, i.e. it has a maximum turning-point.
    - However, there might be further "moments" in your data: throw in an X^3 too, and see what happens.
    - You might come up with some different alternative functional forms. Estimate these, and check the goodness of fit.
    - Never underestimate the powerful eye-ball test: Plot the data and look at it!

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

- **Example 1: Financial satisfaction and income**
  - I have some data on Financial Satisfaction and Income, from wave 14 of the HILDA survey.
  - Financial Satisfaction (FS) is measured on a scale between 0 (totally dissatisfied) and 10 (totally satisfied).
  - I measure income (INC) as annual equivalised disposable household income (why do you think I use household income instead of personal income?)
  - I want to estimate the association between FS and INC, but I am not sure of the correct functional form. I try lin-lin, lin-log, and quadratic:

(1) $FS_i = b_0 + b_1 INC_i + u_i$

(2) $FS_i = b_0 + b_1 INC + b_2 (INC_i)^2 + u_i$

(3) $FS_i = b_0 + b_1 (\ln INC_i) + u_i$

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

- **Example 1: Financial satisfaction and income**

(1) $FS_i = b_0 + b_1 INC_i + u_i$

```
. regress FS INC

      Source |       SS           df       MS      Number of obs   =    17,488
-------------+----------------------------------   F(1, 17486)     =    780.17
       Model |  3735.78915         1  3735.78915   Prob > F        =    0.0000
    Residual |  83730.1486    17,486  4.78841065   R-squared       =    0.0427
-------------+----------------------------------   Adj R-squared   =    0.0427
       Total |  87465.9377    17,487  5.00176918   Root MSE        =    2.1882

------------------------------------------------------------------------------
          FS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         INC |   .0000102   3.65e-07    27.93   0.000     9.48e-06    .0000109
       _cons |   5.939023   .0259705   228.68   0.000     5.888118    5.989927
------------------------------------------------------------------------------
```


Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

- **Example 1: Financial satisfaction and income**

(1) $FS_i = b_0 + b_1 INC_i + u_i$ (room for extra notes)

31

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

- **Example 1: Financial satisfaction and income**

(2) $FS_i = b_0 + b_1 INC + b_2(INC_i)^2 + u_i$

```
. generate INCSQ=INC^2

. regress FS INC INCSQ

      Source |       SS           df       MS      Number of obs   =    17,488
-------------+----------------------------------   F(2, 17485)     =    576.04
       Model |  5406.83368         2  2703.41684   Prob > F        =    0.0000
    Residual |  82059.1041    17,485  4.69311433   R-squared       =    0.0618
-------------+----------------------------------   Adj R-squared   =    0.0617
       Total |  87465.9377    17,487  5.00176918   Root MSE        =    2.1664

------------------------------------------------------------------------------
          FS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         INC |   .0000202   6.44e-07    31.45   0.000      .000019    .0000215
       INCSQ |  -2.63e-11   1.39e-12   -18.87   0.000    -2.90e-11   -2.36e-11
       _cons |   5.520841   .0339438   162.65   0.000     5.454308    5.587374
------------------------------------------------------------------------------
```


32

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

The University of Western Australia
Achieve International Excellence

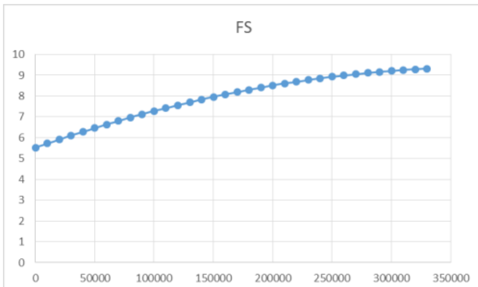
- **Example 1: Financial satisfaction and income**

(2) $FS_i = b_0 + b_1 INC + b_2(INC_i)^2 + u_i$ (room for extra notes)

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Example 1: Financial satisfaction and income** $FS_i = b_0 + b_1 INC + b_2(INC_i)^2 + u_i$
  - You can find the theoretical maximum by using calculus (or the short-cut formula for quadratic functions), and then see whether this maximum lies within a feasible range:
    - X-value for turning point = $-\frac{b_1}{2b_2} =$
    - How many people will have incomes above this value? (52, as it turns out… about 0.3% of the sample)
  - Even better: Plot the model in Excel, using a range of feasible INC values.



Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Example 1: Financial satisfaction and income:**

$$(3)\ FS_i = b_0 + b_1(\ln INC_i) + u_i$$

```
. generate LNINC=ln(INC+1)

. regress FS LNINC

      Source |       SS           df       MS      Number of obs   =    17,488
-------------+----------------------------------   F(1, 17486)     =    695.63
       Model |  3346.47095         1  3346.47095   Prob > F        =    0.0000
    Residual |  84119.4668    17,486  4.81067521   R-squared       =    0.0383
-------------+----------------------------------   Adj R-squared   =    0.0382
       Total |  87465.9377    17,487  5.00176918   Root MSE        =    2.1933

------------------------------------------------------------------------------
          FS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       LNINC |   .5235245   .0198494    26.37   0.000     .4846178    .5624313
       _cons |     .90387   .2127524     4.25   0.000     .4868541    1.320886
------------------------------------------------------------------------------
```

35

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Example 1: Financial satisfaction and income:**

$$(3)\ FS_i = b_0 + b_1(\ln INC_i) + u_i$$

(Room for extra notes)

36

Topic 2: Univariate Regression and OLS
e) Working with nonlinear patterns

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

- **Example 1: Financial satisfaction and income**
  - You can actually also check whether there are departures from linearity in the lin-log model: Throw in a squared log term and see if its coefficient is statistically significant!

$$(4)\ FS_i = b_0 + b_1(\ln INC_i) + b_1(\ln INC_i)^2 + u_i$$

```
. generate LNINCSQ=LNINC^2

. regress FS LNINC LNINCSQ

      Source |       SS           df       MS      Number of obs   =    17,488
-------------+----------------------------------   F(2, 17485)     =    568.92
       Model |  5344.08821         2  2672.04411   Prob > F        =    0.0000
    Residual |  82121.8495    17,485  4.69670286   R-squared       =    0.0611
-------------+----------------------------------   Adj R-squared   =    0.0610
       Total |  87465.9377    17,487  5.00176918   Root MSE        =    2.1672

------------------------------------------------------------------------------
          FS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       LNINC |  -.7656264   .0655139   -11.69   0.000    -.8940402   -.6372127
     LNINCSQ |   .0769897   .0037331    20.62   0.000     .0696724     .084307
       _cons |   5.834583   .3183587    18.33   0.000     5.210568    6.458598
------------------------------------------------------------------------------
```

37

Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

- **Example 1: Financial satisfaction and income**

$$(4)\ FS_i = b_0 + b_1(\ln INC_i) + b_1(\ln INC_i)^2 + u_i$$

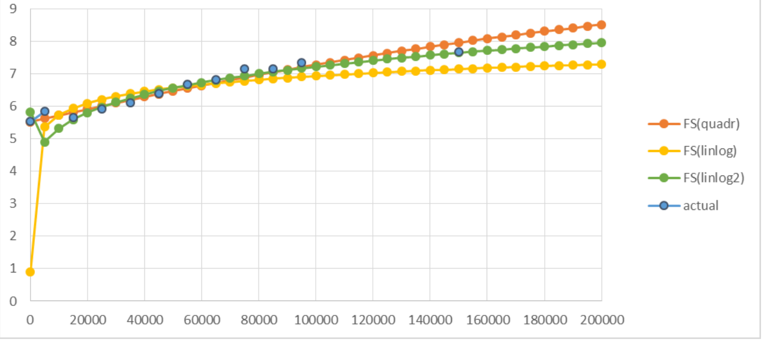
(Room for extra notes)

38

Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

- **Example 1: Financial satisfaction and income**
  - Let's see how they all compare:



- They all perform well across the mid-range of incomes, except for the lin-log model at low incomes!!! Terrible….
- The lin-log model implies a slipe that is too flat across the mid-range. The quadratic specification is better, but over-shoots the mark for higher income levels.
- The squared log model performs best – is most flexible.

Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

- **Example 2: Life satisfaction and age**
  - I have some data on Life Satisfaction and age, from wave 14 of the HILDA survey.
  - Life Satisfaction (LS) is measured on a scale between 0 (totally dissatisfied) and 10 (totally satisfied).
  - I want to estimate the association between LS and age. The literature tells me that happiness is U-shaped across the life span. If so, the correct specification is quadratic:

$$LS_i = b_0 + b_1 AGE_i + b_2\,\mathrm{AGE}_i^2 + u_i$$

```
. regress ls age agesq

      Source |       SS           df       MS      Number of obs   =    17,503
-------------+----------------------------------   F(2, 17500)     =    185.50
       Model |  771.455339         2  385.72767   Prob > F        =    0.0000
    Residual |  36389.6511    17,500  2.07940863   R-squared       =    0.0208
-------------+----------------------------------   Adj R-squared   =    0.0206
       Total |  37161.1064    17,502  2.12324914   Root MSE        =     1.442

------------------------------------------------------------------------------
          ls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0496758   .0028503   -17.43   0.000    -.0552627    -.044089
       agesq |   .0005481   .0000293    18.72   0.000     .0004907    .0006055
       _cons |   8.835998   .0622126   142.03   0.000     8.714055    8.957941
------------------------------------------------------------------------------
```


Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

- **Example 2: Life satisfaction and age**
  - But is there additional nonlinearity in this relationship? Throw in a cubed term…

$$LS_i = b_0 + b_1 AGE_i + b_2\,\mathrm{AGE}_i^2 + b_3\,\mathrm{AGE}_i^3 + u_i$$

```
. regress ls age agesq agecubed

      Source |       SS           df       MS      Number of obs   =    17,503
-------------+----------------------------------   F(3, 17499)     =    144.04
       Model |   895.56878         3  298.522927   Prob > F        =    0.0000
    Residual |  36265.5377    17,499  2.07243486   R-squared       =    0.0241
-------------+----------------------------------   Adj R-squared   =    0.0239
       Total |  37161.1064    17,502  2.12324914   Root MSE        =    1.4396

------------------------------------------------------------------------------
          ls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.1254487   .0101965   -12.30   0.000    -.1454348   -.1054626
       agesq |   .0022349   .0002199    10.16   0.000     .0018038    .0026659
    agecubed |  -.0000112   1.45e-06    -7.74   0.000    -.0000141   -8.39e-06
       _cons |   9.815211   .1409551    69.63   0.000     9.538925     10.0915
------------------------------------------------------------------------------
```

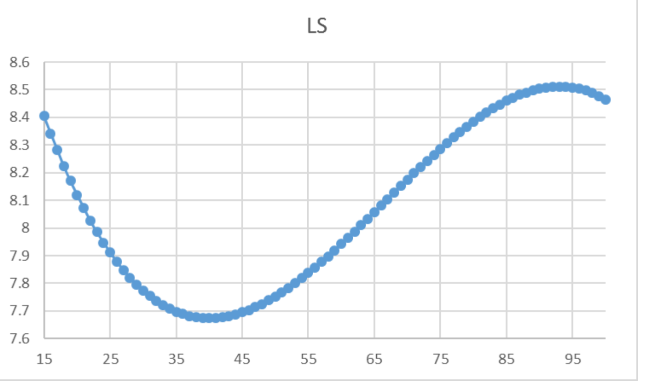
41

Topic 2: Univariate Regression and OLS

e) Working with nonlinear patterns

- **Example 2: Life satisfaction and age**

$$LS_i = b_0 + b_1 AGE_i + b_2\,\mathrm{AGE}_i^2 + b_3\,\mathrm{AGE}_i^3 + u_i$$



Topic 2(ii): Univariate Regression and OLS

Suggested guide to further reading

- **Topic 2 (ii): Univariate Regression:** $Y = b_0 + b_1X + e$

  a) The classical assumptions behind OLS:
  - Wooldridge Ch 4
  - Gujarati Ch 3.2-3.4 (and Ch 4.1-4.3)
  - Kennedy Ch 2.5-2.8

  e) Working with nonlinear relationships:
  - Gujarati Ch 2.3; 6.4-6.8
  - Wooldridge Ch 4.5