EC 32 ECONOMETRICS II

Nikhil Damodaran

Spring 2021

O.P. Jindal Global University
A Private University Promoting Public Service

Lecture 1: 1 Feb 2021: 10:00AM

45 students present, discussed syllabus, assessment and the course structure.

1. *Philosophy of econometrics*: learn econometrics by thinking about the (a) technical details of an estimation and inference process later (b) along with application of the underlying economic idea. Often it is easier to think about the underlying economics and evaluate estimation and inference.

2. *Course overview*: We will cover MLR: its crucial extensions and touch base on issues with estimation of cross section models.

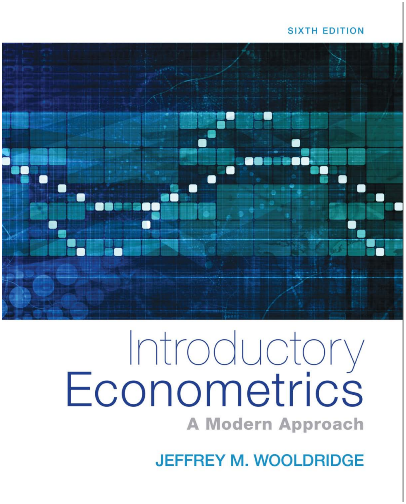
3. *Why choose Panel?*: We discussed how preliminary panel data methods could help us answer newer questions and what is likely to fail in cross sectional methods.

4. *Why skip time series?*: Why we cannot cover time series methods in a second course?

Lecture 2: 3 Feb 2021: 12:30PM
24 students present

- Rescaling data - for x, y or both
- Beta regressions, log transformation
- Forms survey on usage of STATA and R

6 Multiple Regression Analysis: Further Issues



Introduction: This chapter is important from the point of view of an applied economist and covers aspects which would be encountered in the first stage of data wrangling and model design. So be careful and internalize these ideas to succeed in empirical work.

This chapter answers the following questions

1. What happens if I scale my data from thousands to millions, or change the level of the independent or the dependent variable?

2. What happens when either the independent or dependent variable in a regression are transformed to logs or quardratics?

3. How much should I trust R-square for a regression?

4. How to obtain confidence intervals for OLS predictions?

6.1 Effects of Data Scaling on OLS Statistics

1. Often, data scaling is used for cosmetic purposes, such as to reduce the number of zeros after a decimal point in an estimated coefficient.

2. What happens to regression statistics when you scale data?
   Case 1: Changing dependent variable: y

   (a) $\hat{\beta}$ gets scaled down proportionately,

   (b) Statistical significance i.e. t-stat identical

   (c) R-squared or a measure of goodness of fit is also identical.

   (d) Standard error of the regression (SER) changes. This only reflects difference in unit of measurement.

3. **Why**: scaling of regression should not affect the relation if all variables are scaled proportionately.

4. *Example*: Consider the model of infant birth weights (lb vs. ounces) explained by mother's smoking habits, controlling for family income

$$bwght = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 faminc$$

   (a) $bwght$ is child birth weight, in ounces where 1 lb = 16 ounces.
   (b) $cigs$ is number of ciggarettes smoked by the mother while pregnant, per day.
   (c) $faminc$ is annual family income, in thousands of dollars.
   (d) Notice $\hat{\beta}$, calculate t-stat, R-squared and SER.

TABLE 6.1 Effects of Data Scaling

Dependent Variable	(1) *bwght*	(2) *bwghtlbs*	(3) *bwght*
Independent Variables			
cigs	−.4634 (.0916)	−.0289 (.0057)	—
packs	—	—	−9.268 (1.832)
faminc	.0927 (.0292)	.0058 (.0018)	.0927 (.0292)
intercept	116.974 (1.049)	7.3109 (.0656)	116.974 (1.049)
Observations	1,388	1,388	1,388
R-Squared	.0298	.0298	.0298
SSR	557,485.51	2,177.6778	557,485.51
SER	20.063	1.2539	20.063

5. **Case 2: changing scale of independent variable:** x - one cigarette pack i.e. $cigs/20$ (table 3)
   (a) Coefficients change cause we alter the rate of change of x-variable.
   (b) The standard error on *packs* (1.83) is 20 times larger than that on *cigs* (0.0916) in column 3

6. **Log dependent variable**: In the case of $log(x_i)$, there is no change in slope when the data is transformed. Only the intercept is scaled.
   (a) Estimating elasticities in a log-log model of the following form is invariant to scaling data.

$$\widehat{\ln y} = \hat{\beta}_0 + \hat{\beta}_1 \ln x$$

   Here $\hat{\beta}_1$ is the elasticity estimate.

6.1.1 Beta Coefficients

1. **Level differences in y and x variables**: When the level of independent and dependent variables are very different, then its best to ask *what happens when x is one standard deviation higher or lower*.

   (a) *Standardized variables*: A variable is standardized in the sample by subtracting off its mean and dividing by its standard deviation.

   (b) Also know as the z-score transformation.

   (c) Makes different variables comparable.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{u}_i$$

When we standardize using σ_y and σ_1 for x_1 we obtain the following equation in their z-scores:

$$\frac{y_i - \bar{y}}{\hat{\sigma}_y} = \frac{\hat{\sigma}_1}{\hat{\sigma}_y}\hat{\beta}_1 \frac{x_{i1} - \bar{x}}{\hat{\sigma}_1} + \frac{\hat{\sigma}_2}{\hat{\sigma}_y}\hat{\beta}_2 \frac{x_{i2} - \bar{x}}{\hat{\sigma}_2} + \hat{u}_i$$

Thus the new standardized coefficients are a transformation of the original $\hat{\beta}$ and are called "standardized coefficients".

2. **Interpreting new** β: if x_1 increases by one standard deviation, then $\hat{y}$ changes by $\hat{b}_1 = \frac{\hat{\sigma}_1}{\hat{\sigma}_y}\hat{\beta}_1$ standard deviations.

   (a) **Note**: In a standard OLS equation, it is not possible to simply look at the size of different coefficients and conclude that the explanatory variable with the largest coefficient is "the most important."

   (b) Even in elasticity regressions, if the variability in x is small and y is large (or vice versa), you could compute standardized coefficients.

(c) *Example*: What is the standardized effect of pollution on house prices? (HPRICE2)

$$\widehat{zprice} = -0.340 znox - 0.143 zcrime + 0.515 zrooms$$
$$-0.235 zdist - 0.270 zstratio$$

This equation shows that a one standard deviation increase in nox decreases price by .34 standard deviation; a one standard deviation increase in crime reduces price by .14 standard deviation. Thus, the same relative movement of pollution in the population has a larger effect on housing prices than crime does.

6.2 More on Functional Forms

6.2.1 Log Models

1. Interpreting log log models

$$\begin{aligned} log(price) = &\beta_0 + \beta_1 log(nox) + \beta_2 rooms + u \\ & 9.23(0.19) - 0.718(0.066) + 0.306(0.019) \end{aligned}$$

$100\hat{\beta}_2$: semi-elasticity of price with respect to rooms cause only y is in logs. However this is an approximation only for small changes in y i.e. $\Delta log(price)$. To track the exact partial effect in a semi-log model we use the following:

$$\%\Delta\hat{y} = 100 \cdot [exp(\hat{\beta}_2 \Delta\hat{x}_2) - 1]$$

With correction: a unit increase in number of rooms increases house price by 35.8%. (**check**!!) What is the coefficient when number of rooms reduce by one unit?

2. Prevalance of log models in empirical work

   (a) When $y > 0$, models using log(y) as the dependent variable often satisfy the CLM assumptions more closely than models using the level of y.

   (b) Taking the log of a variable often narrows its range hence be less sensitive to outliers.

3. Common rules:

   (a) A variable is a positive dollar amount (price/revenue) or large integer number (GDP/population), the log is often taken

   (b) Variables that are measured in years – such as education, experience, tenure, age, and so on – usually appear in their original form.

       Caution: Remember, if *unem* goes from 8 to 9, this is an increase of one percentage point, but a 12.5% increase from the initial unemployment level. Using the log means that we are looking at the percentage change in the unemployment rate: $log(9) - log(8) \approxeq 0.118$ or 11.8%, which is the logarithmic approximation to the actual 12.5% increase.

   (c) When variable contains few zeros, then use $log(1 + y)$ only when such instances are few (say 2% of total). For other cases use Tobit models

Lecture 3: 8 Feb 2021: 3:30 PM
21 students present

- Quardratic, Interaction and APE from chapter 6
- Homework: get the APE from actual data

6.2.2 Model with Quadratics

1. **Quardratic transformations in linear equations**: are also used quite often in applied economics to capture decreasing or increasing marginal effects and to tease out a non-linear relation between y and the x.

   $$wage = \beta_0 + \beta_1 exper + \beta_2 exper^2$$

   What is the partial effect of $exper$?

   $$\Delta\hat{y} \approxeq (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x$$

   What does this imply? Level of experience matters in calculating the partial effects.

2. *Example*: The previous wage model is estimated to be:

$$\hat{wage} = 3.73 + 0.298exper - 0.0061exper^2$$

Interpretation: when $\beta_2 < 0$ when $\beta_1 > 0$ we know that experience has diminishing effect on wage. A second year of experience only gives an additional 28.6 cents in comparison with the first hour which gives 29.8 cents.

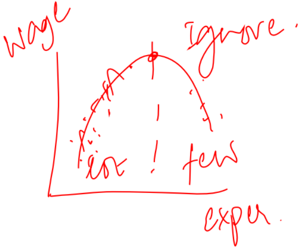
3. What if $\hat{\beta}_1 > 0$ while $\hat{\beta}_2 < 0$. The relation is parabolic and we are then bothered about the turning point (vertex) of the parabola.

   - ▷ Turning point is a value of x obtained from the regression exercise:

$$x^* = \left| \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right|$$

   - ▷ Turning point for experience comes out to be 24.4 years. Does this make sense? How do we interpret the economics underlying this estimate?

4. **First possibility**: If there are very few values on the RHS, then we can make a claim that the estimates are biased by outliers in data.



5. **Second possibility**: the estimated effect of *exper* on *wage* is biased because we have controlled for no other factors, or because the functional relationship between *wage* and *exper* in model equation is not entirely correct

6. What if the estimated β_1 and β_2 for a quardratic model are of same sign for the range of data?

7. Check out the example of house prices to understand application of quardratic effects.

8. Quardratic and log models can be combined to estimate models in which the elasticity is not constant at every point in data.
   Consider the house price data in which we take log-quardratic nitro-oxide as an explanatory variable.

$$\begin{aligned} log(price) &= \beta_0 + \beta_1 log(nox) + \beta_2 [log(nox)]^2 \\ &+ \beta_3 crime + \beta_4 rooms + \beta_5 rooms^2 + \beta_6 stratio + u \end{aligned}$$

6.2.3 Model with Interaction Terms

1. *Example*: Consider a model which tries to explain test scores in a final exam ($stndfnl$) in terms of percentage of classes attended, prior college grade point average, and ACT scores:

$$\begin{aligned} stndfnl =& \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA^2 \\ &+ \beta_5 ACT^2 + \beta_6 priGPA \cdot atndrte + u \end{aligned}$$

   We think that it is the combination of attendance as well as prior sincerity on students which determines their test scores.

2. Regression result:

Using the 680 observations in ATTEND, for students in a course on microeconomic principles, the estimated equation is

$$\begin{aligned}\widehat{stndfnl} &= \underset{(1.36)}{2.05} - \underset{(.0102)}{.0067}\, atndrte - \underset{(.48)}{1.63}\, priGPA - \underset{(.098)}{.128}\, ACT \\ &+ \underset{(.101)}{.296}\, priGPA^2 + \underset{(.0022)}{.0045}\, ACT^2 + \underset{(.0043)}{.0056}\, priGPA \cdot atndrte \qquad [6.19] \\ &n = 680,\ R^2 = .229,\ \bar{R}^2 = .222.\end{aligned}$$

3. **Caution while interpretation**:

- We cannot interpret β_1 in isolation. It does not make sense to say what would be the test score if a student attended the classes if his $priGPA = 0$.
- What are the β_1 and β_6 t-values? What does it tell me about β_1?
- What is the partial effect of atndrte on stndful?

- Partial effect of $atndrte$ on $stndful$ when $pri\bar{G}PA = 2.59$ is $-0.0067+0.0056(2.59) \approxeq 0.0078$: Because $atndrte$ is measured as a percentage, it means that a 10 percentage point increase in $atndrte$ increases $stndfnl$ by .078 standard deviations from the mean final exam score.

- Correct partial effect of $priGPA$

$$\frac{\partial\ stndfnl}{\partial\ priGPA} = -1.63 + 2(0.296)priGPA + 0.0056(atndrte)$$

- *Homework*: use the data in ATTEND to compute the partial effect of priGPA from the data.

6.2.4 Average Partial Effects

1. Model with interactions and quardratics have partial effects which have level variables such as the partial effect of x_2 in the following model:

$$y = \beta_0 + \beta_1 x_1^2 + \beta_2 x_2 + \beta_3 x_1 x_2$$

2. **Problem**: we do not have one number from regression estiamtes which give us the partial effect.

3. **Potential solution**: We compute partial effects across each unit of the sample and take its average – called the Average Partial Effect

$$\text{Partial Effect} : \frac{\partial y}{\partial x_2} = \beta_2 + \beta_3 x_1$$

$$\text{Average Partial Effect} := \beta_2 + \beta_3 \frac{1}{n} \sum_{i=0}^{n} x_{1i}$$

Lecture 4: 10 Feb 2021
25 students present

- Target: remaining chapter 6
  - What does R^2 measure? Can we rely on it entirely?
  - What is the adjustment in $\bar{R}^2$? How do we use it and what are its limitations?
  - Choosing between non-nested models using $\bar{R}^2$
  - Over-controlling using the beer example

6.3 More on Goodness-of-Fit and Selection of Regressors

1. What is R-squared? (Chapter 3)
   It is a ratio of the explained sum of squares (SSE) to total sum of squares (SST) or could be expressed as a measure of what is potentially left unexplained through the residual sum of squares (SSR)

$$SST = \sum_{i=0}^{n} (y_i - \bar{y})^2 \qquad SSE = \sum_{i=0}^{n} (\hat{y}_i - \bar{y})^2$$

$$SSR = \sum_{i=0}^{n} \hat{u}_i^2 \qquad R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

   Interpreted as the proportion of the sample variation in y_i that is explained by the OLS regression line

2. Caution: But it has nothing to do with baised-ness of the estimates or the accuracy of the functional form.

   - Could potentially increase sample size and reduce error variance, increasing the precision of the estimates. But the model could still be faulty!
   - Could have a variable whose true explanatory power (in population) is only 25% so R-squared will not go above 25% no matter how large the sample is!
   - With a carefully designed random experiment, you could potentially obtain a $\hat{\beta}$ which is true to its population counterpart, despite a very low R^2

3. Low R-suqred will make prediction difficult. If our estimates are precise through random experimental data, but they only explain 10% of total dependent variable behavior, then we cannot predict accurately.

6.3.1 Adjusted R-Squared

1. Adjusted R-square: is nothing but R-squared adjusted for the size of the data (n) and the number of independent variables (k) in the model

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$$

2. Note: $\bar{R}^2$ is not generally known to be a better estimator than R^2. Then why do we use $\bar{R}^2$?

3. $\bar{R}^2$ tell us how many x's are too many x's?

   - Since adding independent variables increases explanatory power of the model, should we go on adding x's?

     No, since you gain explanatory power (perhaps), but you loose degree of freedom which is reflected in $\bar{R}^2$.

   - How do we know which addition is good?

     If we add a new independent variable to a regression equation, $\bar{R}^2$ increases if, and only if, the t statistic on the new variable is greater than one in absolute value.

4. Relation between $\bar{R}^2$ and R^2

$$\bar{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

6.3.2 Choosing between Non-nested models

1. Remember: usage of F-stat for comparing strength of nested models (chapter 4)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$
$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + u$$

   Second model is obtained when you estimate the first model under the restriction $\beta_2 = 0$. These are called nested models. (Do exercise C5 in chapter 4: next week Monday)

2. Example of non nested models: consider the problem of whether an increase in sales causes firms to invest more in research and development as follows:

$$rdintens = \beta_0 + \beta_1 log(sales) + u \qquad R^2 = 0.061 \quad \bar{R}^2 = 0.030$$
$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2 + u \qquad R^2 = 0.148 \quad \bar{R}^2 = 0.090$$

   Strange example: because the second model is better by looking at both the measures.

3. **Bottom line**:

    (a) when you suspect how many x's, then use $\bar{R}^2$

    (b) $\bar{R}^2$ cannot help you choose the type of functional form - total variation in y and $log(y)$ are not same

    (c) most likely, $\bar{R}^2$ might just be a double check rather than your go to measure as you would not have a lot of independent variables in most applied work.

6.3.3 Over-controlling and Adding Too Much

1. Consider the following example where we want to explain traffic fatalities (fatilities) as a function of tax on alcohol (tax) controlling for other co-variates. We do not use the following model:

$$\begin{aligned} fatalities = &\beta_0 + \beta_1 tax + \beta_2 beercons + \beta_3 miles \\ &+ \beta_4 percmale + \beta_5 perc16-21 \end{aligned}$$

   Why is this wrong: β_1 measures the difference in fatalities due to a one percentage point increase in tax, holding *beercons* fixed.

2. **Over-controlling**: We want the effect of tax on fatalities and not tax on beer consumption itself.

   (a) Potentially unsure of how the effects of the independent variable are affecting the dependent variable – overcontrolling by accident

   (b) Potentially unsure of isolating effect of desired independent variable of interest – overcontrolling on intent

3. Caution: If we remember that different models serve different purposes, and we focus on the ceteris paribus interpretation of regression, then we will not include the wrong factors in a regression model.

   (a) Add variables when they are sure to affect y and are uncorrelated with all other x's – no harm, but how much does it contribute?

      ▷ Read the beer explanation (page 200)

   (b) Adding extra x's will not ensure an unbiased estimation, but might just ensure more precise estimates

   (c) We need not worry about whether some of our explanatory variables are "endogenous"—provided these variables themselves are not affected by the policy

Lecture 5: 15 Feb 2021

26 students present

- Prediction interval: How do we get an estimate of the $\hat{y}$ for particular values of $\mathbf{x}$?
- How do we build confidence interval for $\hat{y}$ for any value of $\mathbf{x}$ in the population?
- Residual analysis
- Predicting $\hat{y}$ when you run a model for $\widehat{log}(y)$
    1. Smearing estimate
    2. Estimate through origin
- Bootstrap standard errors

6.4 Prediction and Residual Analysis

6.4.1 Prediction Error and Interval

1. Consider the estimated model:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

2. For $c_1, c_2, \cdots, c_k$, a data point, the predicted value of $\hat{y}$ is

$$\hat{\theta} = E(y|x_1 = c_1, x_2 = c_2, \ldots, x_k = c_k)$$

   To get a confidence interval on this estimate, we need a way of calculating its standard error.

(a) Run this model instead:

$$y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + \cdots + \beta_k(x_k - c_k) + u$$

(b) Estimate of $\hat{\theta}$ will be $\hat{\theta}_0$ and its CI can be obtained using its standard error.

(c) Note: when $c_1 = \bar{x}_1$ and so on, you get the smallest SE and θ_0 merely measures the part of variation in dependent variable in your sample which you could explain.

3. Now consider a value in data y^0 *but not in our sample.* In order to obtain a prediction error (out of sample error) we calculate what our regression predicts its value would be and what its value actually is:

$$\hat{e}^0 = y^0 - \hat{y}^0$$

4. Prediction error will have a standard deviation, which gives us the CI for our perdiction. The **variance of our prediction error** is :

$$Var(\hat{e}^0) = Var(\hat{y}^0) + Var(u^0)$$

where $Var(\hat{y}^0)$ is the sample variance in your data and depends upon how large your sample is.

5. $Var(u^0)$ does not depend on your sample and is unobservable σ^2 However, we can use sample estimate of unobservable σ^2 to obtain

$$se(\hat{e}_0) = \left\{ [se(\hat{y}^0)]^2 + \hat{\sigma}^2 \right\}^{\frac{1}{2}}$$

$se(\hat{e}_0)$ has t-distribution with $n - k - 1$ degrees of freedom. This gives us the prediction interval as:

$$\hat{y}^0 \pm t_{0.025} \cdot se(\hat{e}_0)$$

where $t_{0.025} \approx 1.96$ and the above CI gives a 95% interval.

6. Example: Consider a model which tries to explain $colgpa$ which is the cumulative grade point average, $hsize$ is the size of the class graduating, $hsperc$ is academic percentile in graduating class, sat is combined SAT scores, $female$ is gender binary and $athlete$ is a participation binary from GPA2:

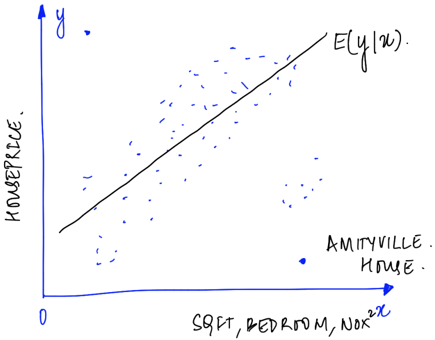
$$\begin{aligned} colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat \\ + \beta_5 female + \beta_6 athlete + u \end{aligned}$$

   (a) Obtain a 95% confidence interval for the expected college GPA when sat = 1200, hsperc = 30 and hsize = 5 (try another combination too!)

   (b) Also obtain a 95% confidence interval for any value of sat, hsperc and hsize

6.4.2 Residual Analysis

1. **Residual analysis**: is when you look at the residual from a regression line to understand observations which are not fit well by the regression i.e. observations away from the line of fit.



You can live in the 'Amityville Horror' home for $850,000

Published: July 2, 2016 at 10:55 a.m. ET

—

By Daniel Goldstein

112

Allegedly haunted Long Island house was the scene of grisly mass murder in 1974

2. Example: Using the data in HPRICE1, we run a regression of *price* on *lotsize*, *sqrft*, and *bdrms* we uncovered the ghost house

   - **n**: sample of 88 homes
   - **Residuals**: the most negative residual is 2120.206, for the 81st house
   - **price**: the asking price for this house is $120,206 below its predicted price

6.4.3 Predicting y when using log(y)

1. **Log transformations**: typically bring large values closer to down and inflate very small values. Thus it changes the standard deviation σ of the variable which is being transformed.

2. Consider the semi-log model:

$$log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

If the model is estimated correctly, then

$$E(y|x) = e^{\sigma^2/2} exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

3. Interpreting level y: implies adjusting for standard error from log models for all the parameters. We can simply use sample standard error $\hat{\sigma}^2$ as proxy for σ^2.

   - This prediction of level y based on the correction is consistent relying on normally distributed errors
   - There are no unbiased estimates.

4. **Correct prediction of level without normality of errors**:

$$E(y|x) = \alpha_0 exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

where α_0 is the expected value of $exp(u)$ or $\alpha_0 = E[exp(u)]$. How do we estimated α_0?

   (a) **Smearing estimate**: estimated the $log(y)$ model with sample data and then use the OLS residual $\hat{u}_i$ to estimate $\hat{\alpha_0}$ as

$$\hat{\alpha_0} = \frac{1}{n} \sum_{i=0}^{n} exp(\hat{u}_i)$$

   (b) **Estimate through origin**: Define $m_i = exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$ and regress m_i on $exp(log(y_i))$ without intercept. The β_j hence obtained is an estimate of α_0 and we denote it by $\check{\alpha_0}$

5. **Goodness of fit with $log(y)$ model**: For the level model estimated as follows:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

R^2 measures squared correlation between y_i and $\hat{y}_i$. But what we have is $\hat{y}_i = \hat{\alpha_0} m_i$. We can use any measure of $\hat{\alpha_0}$ to obtain R^2 because correlations are not affected by change of scale through multiplication.

6. **Bottom line**: use the R^2 compatible with your estimate of $\hat{\alpha_0}$

6.4.4 Bootstrap Standard Errors

1. **Idea of Bootstrapping**: Imagine that you have a sample (large enough), but you have no idea of what the sampling distribution looks like. At times the best proxy of sampling distribution is to treat the big sample as population and draw samples from it.

   (a) Such methods are called re-sampling methods

   (b) Suppose we have an estimate $\hat{\theta}$ which is an estimate of θ. To obtain CIs for $\hat{\theta}$ we do re-sampling called bootstrapping.

2. Bootstrapping original sample of size n by drawing samples with replacement of size n and estimating θ using these different samples. Let us assume we generate m samples. Then the bootstrap standard error of $\hat{\theta}$ is

$$bse(\hat{\theta}) = \left[\frac{1}{(m-1)} \sum_{b=1}^{m} (\hat{\theta}^b - \bar{\hat{\theta}})^2\right],$$

where $\bar{\hat{\theta}}$ is the average of the bootstrap estimates.