Lecture 18

Confidence Intervals for Predictions

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 X_i + u_i \quad , \quad i = 1, 2, \cdots, N.$$

1. Suppose we are interested in predicting the expected value of y given $X = X_0$, that is,

$$E(y \mid X = X_0) = \beta_0 + \beta_1 X_0 = \mu_{y|X_0}$$

The unbiased predictor of $E(y \mid X = X_0)$ is

$$\widehat{E(y \mid X = X_0)} = \hat{\beta}_0 + \hat{\beta}_1 X_0 .$$

But what is the variance of this prediction?

$$Var\left[\widehat{E(y \mid X = X_0)}\right] = Var(\hat{\beta}_0 + \hat{\beta}_1 X_0)$$

$$= Var(\hat{\beta}_0) + X_0^2 \, Var(\hat{\beta}_1) + 2X_0 Cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$= \frac{\sigma^2 \sum_1^N X_i^2}{N \sum_1^N (X_i - \bar{X})^2} + X_0^2 \frac{\sigma^2}{\sum_1^N (X_i - \bar{X})^2} - 2 X_0 \bar{X} \frac{\sigma^2}{\sum_1^N (X_i - \bar{X})^2}$$

$$= \frac{\sigma^2 \left(\frac{1}{N} \sum_1^N X_i^2 + X_0^2 - 2 X_0 \bar{X} \right)}{\sum_1^N (X_i - \bar{X})^2}$$

$$= \frac{\sigma^2 \left[\frac{1}{N} \sum_1^N (X_i - \bar{X})^2 + (X_0 - \bar{X})^2 \right]}{\sum_1^N (X_i - \bar{X})^2}$$

$$= \sigma^2 \left(\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_1^N (X_i - \bar{X})^2} \right)$$

we might abbreviate our notation somewhat by denoting

$$\sigma^2_{(\hat{\mu_y}|X_0)} = \sigma^2 \left(\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_1^N (X_i - \bar{X})^2} \right).$$

That is,

That is the variance of $\hat{\beta}_0 + \hat{\beta}_1 X_0$ in predicting $E(y|X=X_0) = \beta_0 + \beta_1 X_0$ is denoted $\sigma^2_{y|x}$.

As we can see, the accuracy in predicting $E(y|X=X_0)$ increases as N increases but decreases as the distance between X_0 and $\bar{X}$ increases. (It goes without say, the larger the error variance, σ^2, the less precise the prediction.)

The standard error of the prediction of $E(y|X=X_0)$ is given by

$$se(\hat{\mu}_{y|x_0}) = \hat{\sigma}\left(\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_1^N (X_i - \bar{X})^2}\right)^{1/2}$$

where $\hat{\sigma} = \sqrt{\frac{SSR}{N-k-1}}$.

Now if we assume normality of the errors u_i in the regression model, we can invert the below t-statistic to get a $(1-\alpha)\%$ confidence interval for $E(Y|X=X_0) = \mu_{Y|X_0}$. Note that the random variable

$$t_v = \frac{\hat{\beta}_0 + \hat{\beta}_1 X_0 - E(Y|X=X_0)}{se(\hat{\mu_Y|X_0})}$$

has a t-distribution with $v = N-k-1$ degrees of freedom. Thus,

$$Pr\left(-t_{v,\alpha/2} < \frac{\hat{\beta}_0 + \hat{\beta}_1 X_0 - E(Y|X=X_0)}{se(\hat{\mu_Y|X_0})} < t_{v,\alpha/2}\right) = 1-\alpha$$

Inverting the above expression gives us the $(1-\alpha)\%$ confidence interval for $E(Y|X=X_0)$, namely

$$Pr\left(\hat{\beta}_0 + \hat{\beta}_1 X_0 - t_{v,\alpha/2}\cdot se(\hat{\mu_Y|X_0}) < E(Y|X=X_0) < \hat{\beta}_0 + \hat{\beta}_1 X_0 + t_{v,\alpha/2}\cdot se(\hat{\mu_Y|X_0})\right) = 1-\alpha.$$

Obviously, given the estimated variance-covariance matrix of the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, we can get the standard error of the predictor $\hat{\beta}_0 + \hat{\beta}_1 X_0$ in predicting $E(y|X=X_0) = \beta_0 + \beta_1 X_0$ as

$$se(\hat{\mu}_{y|x_0}) = \Big[\hat{Var}(\hat{\beta}_0) + X_0^2 \, \hat{Var}(\hat{\beta}_1) + 2 X_0 \, \hat{cov}(\hat{\beta}_0, \hat{\beta}_1) \Big]^{1/2}.$$

But is there a way of tricking the computer into getting $se(\hat{\mu}_{y|x_0})$ directly? The answer is yes! Let

$$\theta_0 = \beta_0 + \beta_1 X_0 .$$

Then rewrite the original simple linear regression model as

$$\begin{aligned} y &= \beta_0 + \beta_1 X - \theta_0 + \theta_0 + u \\ &= \theta_0 + \beta_1 (X - X_0) + u \qquad (1). \end{aligned}$$

Estimating equation (1) by OLS we have

$$\hat{y} = \hat{\theta}_0 + \hat{\beta}_1 (X - X_0) \qquad (2)$$

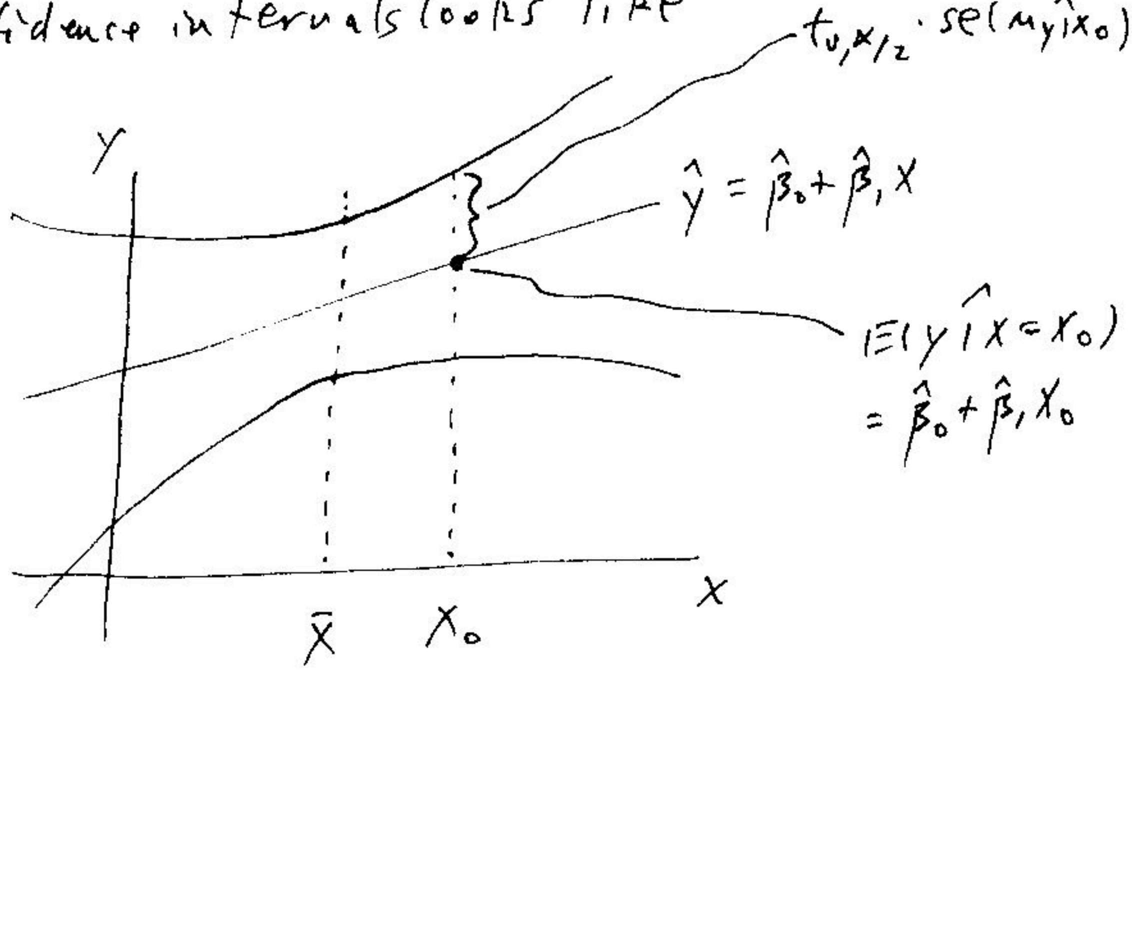
Then in (2) $\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0 = E(y \mid X = X_0)$

as we desire and $se(\hat{\theta}_0) = se(\hat{\mu}_{y|x_0})$ that

we need for forming the $(1-\alpha)\%$ confidence

interval for the prediction of $E(y \mid X = X_0)$. The

confidence intervals looks like



2. Consider the problem of predicting an individual observation on y, say y_0, when $X = X_0$.

The prediction error is then

$$\hat{e}_0 = (y_0 - \hat{y}_0)$$

$$= (y_0 - \hat{\beta}_1 - \hat{\beta}_1 X_0)$$

$$= (\beta_0 + \beta_1 X_0 + u_0 - \hat{\beta}_1 - \hat{\beta}_1 X_0)$$

$$= (\beta_0 + \beta_1 X_0 - \hat{\beta}_1 - \hat{\beta}_1 X_0) + u_0$$

$$= v_0 + u_0 .$$

note: $\hat{\beta}_1 + \hat{\beta}_1 X_0$ is an unbiased predictor of y_0.

But $Var(\hat{e}_0) = Var(v_0) + Var(u_0) + 2\,Cov(v_0, u_0)$

$$= Var(v_0) + Var(u_0)$$

since $Cov(v_0, u_0) = 0$, This follows since $\hat{\beta}_0$ and $\hat{\beta}_1$ are independent of u_0. Now $Var(v_0) = Var(\widehat{E}(y|X = X_0))$

and we have

$$Var(\hat{y}_0) = Var(\hat{e}_0) = Var(\widehat{E}(y|X = X_0)) + \sigma^2$$

$$= \sigma^2 \left[1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_1^N (X_i - \bar{X})^2} \right]$$

Notice the extra term!

The corresponding standard error in predicting y_0 is then

$$\sqrt{\hat{var}(\hat{y}_0)} = \sqrt{\hat{Var}(E(y|X=X_0)) + \hat{\sigma}^2} \equiv se(\hat{y}_0)$$

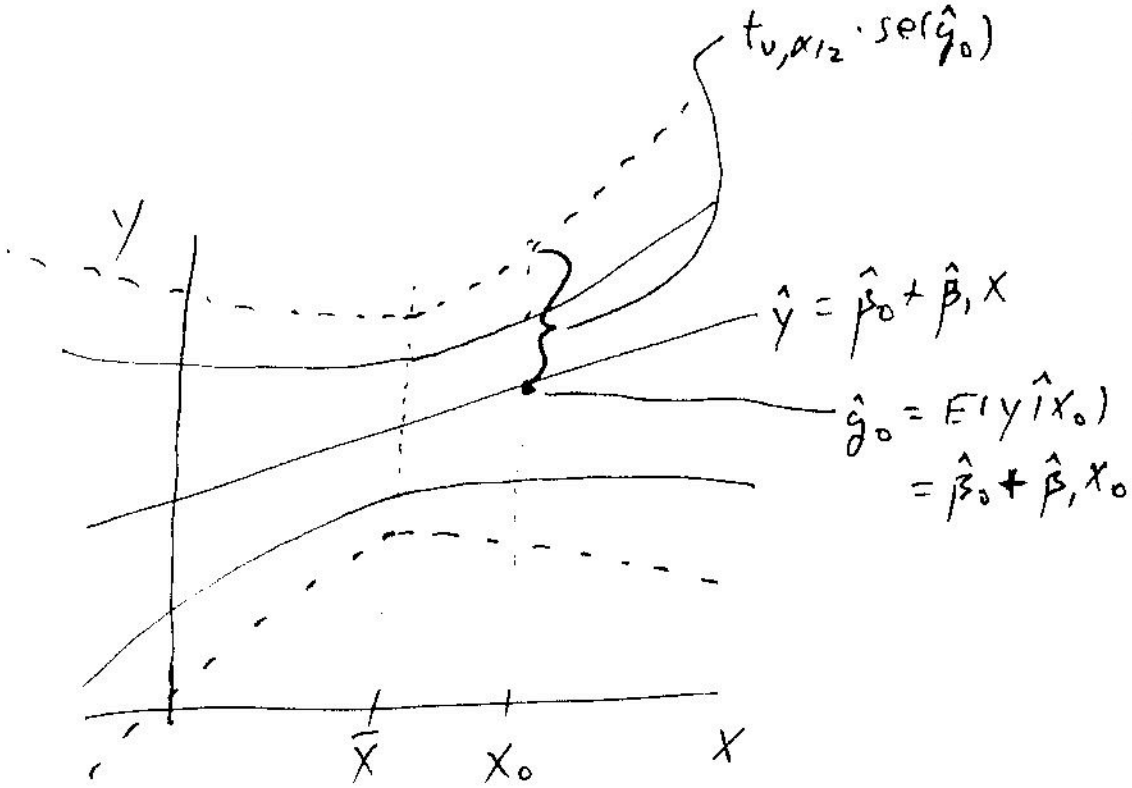
we can get $\hat{Var}(E(y|X=X_0)) = [se(\hat{\theta}_0)]^2$ as previously derived from (2) on page 6.

Also $\hat{\sigma}^2 = SSR/(N-k-1)$ or (standard error of regression)^2. It follows that the $(1-\alpha)\%$ confidence for predicting y_0 when $X = X_0$ is given by

$$pr(\hat{y}_0 - t_{v,\alpha/2} se(\hat{y}_0) < y_0 < \hat{y}_0 + t_{v,\alpha/2} \cdot se(\hat{y}_0)) = 1-\alpha.$$

Notice that the $(1-\alpha)\%$ confidence intervals for predicting y_0 are wider than the corresponding confidence intervals for predicting $E(y|X=X_0)$.

This is represented in the following graph



solid lines = $(1-\alpha)\%$ confidence intervals for $E(y|X=X_0)$

dashed lines = $(1-\alpha)\%$ confidence intervals for y_0

The bottom line is that it is more difficult to predict y_0 than $E(y|X=X_0)$.

3. Suppose that we do not know X_0 but instead have an estimate of X_0, say $\hat{X}_0$. Let

$$\hat{X}_0 = X_0 + \varepsilon$$

where ε is a zero mean error and independent of the regression error u.

The first point to note is that

$$\tilde{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_0$$

is unbiased ~~forecast~~ predictor of y_0. This follows because the forecast error $\tilde{e}_0 = \tilde{y}_0 - y_0$ has a zero expectation.

$$\begin{aligned}
E(\tilde{y}_0 - y_0) &= E(\hat{\beta}_0 + \hat{\beta}_1 \hat{X}_0 - \beta_0 - \beta_1 X_0 - u_0) \\
&= E(\hat{\beta}_0 - \beta_0) + E(\hat{\beta}_1(\hat{X}_0 - X_0)) - E(u_0) \\
&= 0 + E(\hat{\beta}_1 \varepsilon) - 0 \\
&= E(\hat{\beta}_1) E(\varepsilon) = \beta_1 \cdot 0 = 0 .
\end{aligned}$$

Here we have used the fact that $\hat{\beta}_1$ and ε are independent random variables.

Deriving the variance of the forecast error $\tilde{e}_0$ is more complex. Note that

$$E(\tilde{e}_0^2) = E(\tilde{y}_0 - y_0)^2 = E(\hat{\beta}_0 + \hat{\beta}_1 \hat{X}_0 - \beta_0 - \beta_1 X_0 - u_0)^2$$

$$= E[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 \hat{X}_0 - \beta_1 X_0) - u_0]^2$$

$$= E(\hat{\beta}_0 - \beta_0)^2 + E(\hat{\beta}_1 \hat{X}_0 - \beta_1 X_0)^2 + E(u_0^2)$$

$$+ 2E(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 \hat{X}_0 - \beta_1 X_0) \qquad (3)$$

First let us focus on the second term of (3).

$$E(\hat{\beta}_1 \hat{X}_0 - \beta_1 X_0)^2 = E[\hat{\beta}_1(\hat{X}_0 - X_0) + X_0(\hat{\beta}_1 - \beta_1)]^2$$

(since $\hat{\beta}_1 \hat{X}_0 - \beta_1 X_0 = \hat{\beta}_1(\hat{X}_0 - X_0) + X_0(\hat{\beta}_1 - \beta_1)$)

$$= E[\hat{\beta}_1^2(\hat{X}_0 - X_0)^2] + X_0^2 E(\hat{\beta}_1 - \beta_1)^2$$

$$+ 2E[\hat{\beta}_1(\hat{X}_0 - X_0) X_0(\hat{\beta}_1 - \beta_1)]$$

$$= [\beta^2 + Var(\hat{\beta}_1)]\sigma_\varepsilon^2 + X_0^2 Var(\hat{\beta}_1)$$

where $E[\hat{\beta}_1(\hat{X}_0 - X_0) X_0(\hat{\beta}_1 - \beta_1)] = E[\hat{\beta}_1 X_0(\hat{\beta}_1 - \beta_1)\varepsilon]$

$$= E[\hat{\beta}_1 X_0(\hat{\beta}_1 - \beta_1)] \cdot E(\varepsilon) = 0 \text{ and}$$

$$E[\hat{\beta}_1^2(\hat{X}_0 - X_0)^2] = E(\hat{\beta}_1^2 \varepsilon^2) = E(\hat{\beta}_1^2)E(\varepsilon^2)$$

$$= (Var(\hat{\beta}_1) + \beta_1^2)\,\sigma_\varepsilon^2$$

(since $Var(\hat{\beta}_1) = E(\hat{\beta}_1^2) - (E(\hat{\beta}_1))^2 = E(\hat{\beta}_1^2) - \beta_1^2$)

Now Let us turn to the last term of (3)

$$E(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1\hat{X}_0 - \beta_1 X_0)$$

$$= E[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1(\hat{X}_0 - X_0) + X_0(\hat{\beta}_1 - \beta_1))]$$

$$= E(\hat{\beta}_0 - \beta_0)\hat{\beta}_1 \varepsilon + X_0 E(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)$$

$$= E(\hat{\beta}_0 - \beta_0)\hat{\beta}_1 E(\varepsilon) + X_0 \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= X_0 \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

Therefore

$$\sigma_{\tilde{y}_0}^2 \equiv E(\tilde{e}_0^2)$$

$$= Var(\hat{\beta}_0) + [\beta_1^2 + Var(\hat{\beta}_1)]\sigma_\varepsilon^2 + X_0^2 Var(\hat{\beta}_1)$$

$$+ \sigma^2 + 2X_0 \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

In terms of our least-squares estimators, the formula of the ~~error~~ variance of the error in predicting y_0 by means of $\hat{\beta}_0 + \hat{\beta}_1\hat{X}_0$ is

$$\sigma^2_{\tilde{y}_0} = \sigma^2 \left[1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2 + \sigma_\varepsilon^2}{\sum_1^N (X_0 - \bar{X})^2} \right] + \beta_1^2 \sigma_\varepsilon^2$$

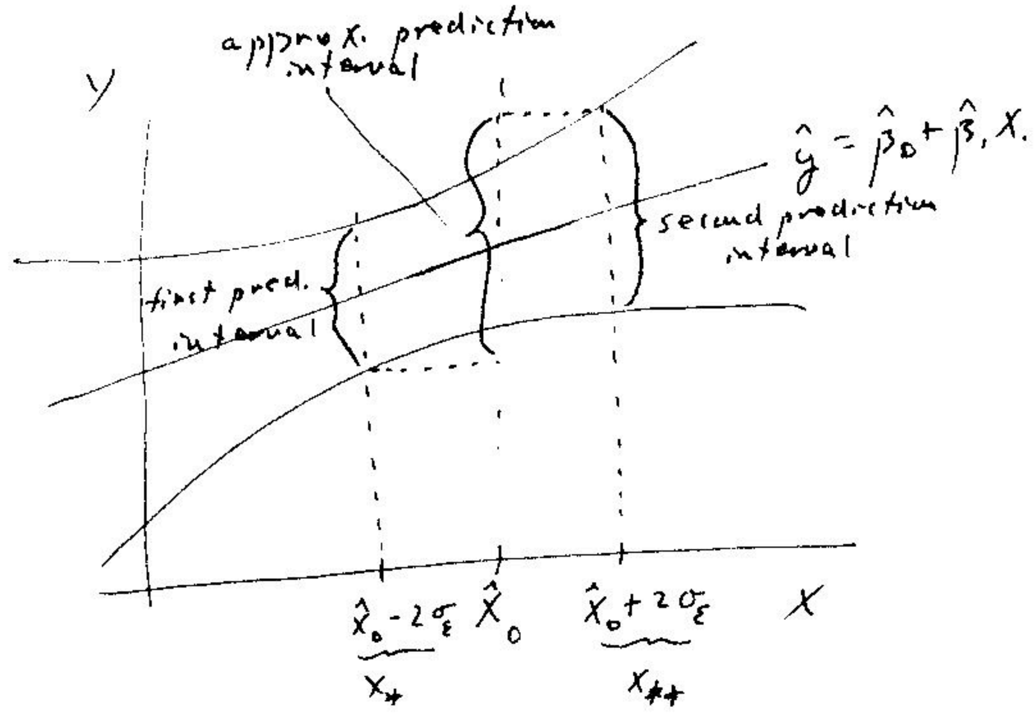
This last formula makes it obvious that when X_0 itself is uncertain, the variance of the forecast error is increased. Therefore, the $(1-\alpha)$% confidence ~~unfortunately~~ intervals associated with the ~~forecasted~~ predicted value $\tilde{y}_0$ will be larger than similar confidence intervals when X_0 is known. Unfortunately, it is quite difficult to describe the confidence intervals for y_0 using $\tilde{y}_0$ because $\tilde{y}_0$ is not normally distributed since it ($\tilde{y}_0$) involves the sum of products of normally distributed variables that is not normally distributed. Therefore, we cannot invert a t-random variable to get a prediction confidence interval for y_0 as we previously have done for the previous cases.

Instead we have to use an approximate solution as recommend by Robert Pindyck and Daniel L. Rubinfeld in their book <u>Econometric Models and Economic Forecasts</u> (New York: McGraw-Hill), 2nd edition, 1981, p. 223.

1. Calculate the $(1-\alpha)\%$ confidence intervals associated with the forecasts that would be obtained were we to select X_0 to be two standard deviations ($2\sigma_\varepsilon$) higher than $\hat{X}_0$, namely $X_* = \hat{X}_0 + 2\sigma_\varepsilon$, and two standard deviations lower than $\hat{X}_0$, namely $X_{**} = \hat{X}_0 - 2\sigma_\varepsilon$, i.e. confidence intervals associated with $\hat{y}_* = \hat{\beta}_0 + \hat{\beta}_1 (X_*)$ and $\hat{y}_{**} = \hat{\beta}_0 + \hat{\beta}_1 (X_{**})$

2. The final prediction interval is taken to be the <u>union</u> of the two confidence intervals; i.e. it contains all the values of $\hat{y}$ common to both confidence intervals.

A graphical representation of this procedure is presented below:



(Wooldridge, pp. 204-206)

See examples 6.5 and 6.6 for examples of constructing prediction intervals for $E(y|X=X_0)$ and y_0 in a multiple regression problem. Also see the SAS program gpa2.sas that is posted on the course website.

Residual Analysis

See pp. 206-207 for a discussion of how residuals from estimated regression models can be used to identify under-valued assets (see the house price example) or law schools that provide the highest value added.

Predicting y when log(y) is the Dependent Variable

Consider the model

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

We can predict log(y) by

$$\widehat{\log(y)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k .$$

To get a predicted value of y we can use the formula:

$$\hat{y} = \exp(\hat{\sigma}/2)\exp(\widehat{\log(y)})$$

where $\hat{\sigma} = \sqrt{\frac{SSR}{N-k-1}}$. For an example of how this formula is utilized see the SAS program ceosal2.sas that is on the website for the course.

Choosing between two regression models, one with y as the dependent variable and the other having the dependent variable log(y).

Consider the two competing models:

$$y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u \qquad (4)$$

$$\log(y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u \qquad (5)$$

First calculate the R^2 of model (4) by computing $R_y^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$.

Second, apply OLS to equation (5) and get the fitted values

$$\widehat{\log}(y) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k .$$

Then convert these log(y) fitted values to fitted values of y by using the formula

$$\tilde{y} = \exp(\hat{\sigma}/2)\exp(\widehat{\log}(y)).$$

Finally calculate the R^2 of the log model as

$$R^2_{\log(y)} = [\mathrm{corr}(y,\tilde{y})]^2$$
$$= \left[\frac{\sum_1^N (y_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}})}{\sqrt{\sum_1^N (y_i-\bar{y})^2 \sum_1^N (y_i - \bar{\tilde{y}})^2}} \right]^2$$
$$= \frac{\left(\sum_1^N (y_i-\bar{y})(\tilde{y}_i - \bar{\tilde{y}})\right)^2}{\sum_1^N (y_i - \bar{y})^2 \sum_1^N (\tilde{y}_i - \bar{\tilde{y}})^2}$$

Then choose the log model if $R^2_{\log(y)} > R^2_y$.

otherwise, choose the model with y as the dependent variable if $R^2_{\log(y)} < R^2_y$.

Alternatively $R^2_{log(y)}$ could be computed as

$$R^2_{log(y)} = 1 - \frac{SSR_{log(y)}}{SST}$$

where $SSR_{log(y)} = \sum_1^N (y_i - \tilde{y}_i)^2$ and

$SST = \sum_1^N (y_i - \bar{y})^2$.

Actually, in this exercise the regressors in eqs. (4) and (5) need not be exactly the same variables but you do have to have the same number of variables per equation to make the comparison. See the SAS program ceo2.sas for a comparison of the regression models of examples (6.7) and (6.8) on pp. 209-210 in the Wooldridge textbook.