

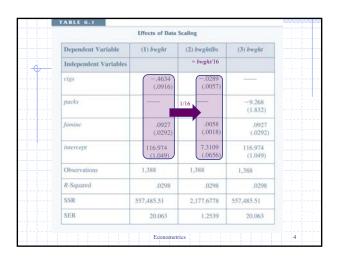
Ch.6 Multiple Regression: Further Issues 1. Effects of Data Scaling on OLS Statistics 2. More on Functional Form 3. More on Goodness-of-Fit & Selection of Regressors 4. Prediction & Residual Analysis*

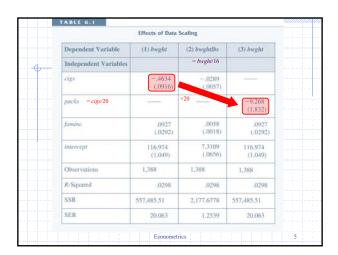
6.1 Effects of Data Scaling

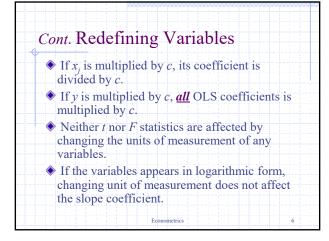
Redefining Variables

Changing the scale of the variable will lead to a corresponding change in the scale of the coefficients and standard errors, but no change in the significance or interpretation.

→ see table 6.1.







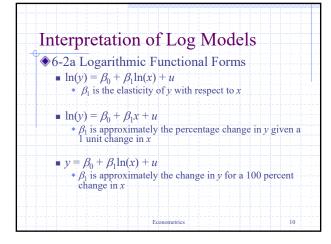
Effects of	Data S	caling	
Dependent Independent	y	cy	у
x_1	$\beta_1 (se_1)$	$c\beta_1 (c*se_1)$	
dx_1			$\beta_1/d (se_1/d)$
x_2	$\beta_2 (se_2)$	$c\beta_2 (c*se_2)$	$\beta_2 (se_2)$
Intercept	$\beta_0 (se_0)$	$c\beta_0 (c*se_0)$	$\beta_0 (se_0)$
R-squared	R^2	R^2	R^2
SSR	SSR	c ² *SSR	SSR
Standard errors in p	parentheses Econon	netrics	

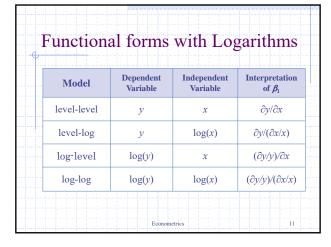
Beta Coefficients

- ♦ Idea is to replace *y* and each *x* variable with a standardized version subtract mean and divide by standard deviation.
- Coefficient reflects standard deviation of *y* for a one standard deviation change in *x*.
 - We can compare the magnitudes of the resulting beta coefficients and conclude that "which variable is most important," etc.
 - Whether we use standardized or unstandardized variables does not affect statistical significance.

Econometrics

6.2 More on Functional Form ◆ Functional Forms OLS can be used for relationships that are not strictly linear in x and y by using nonlinear functions of x and y – will still be linear in the parameters. • the natural log of x, y or both • quadratic forms of x • interactions of x variables





Log models are invariant to the scale of the variables since measuring percent changes. They give a direct estimate of elasticity. For models with y > 0, the conditional distribution is often heteroskedastic or skewed, while ln(y) is much less so. The distribution of ln(y) is more narrow, limiting the effect of outliers.

Why use log models?

Some Rules of Thumb for "log"

- What types of variables are often used in log form?
 - Dollar amounts that must be positive
 - Very large variables, such as population
- What types of variables are often used in level form?
 - Variables measured in years
 - Variables that are a proportion or percent

Econometrics

tries

Quadratic Models

- 6-2b Models with Quadratics
 - For a model of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u,$$

we can't interpret β_1 alone as measuring the change in y with respect to x, we need to take into account β_2 as well, since

$$\frac{\partial \hat{y}}{\partial x} \approx \hat{\beta}_1 + 2\hat{\beta}_2 x$$

conometrice

More on Quadratic Models

- For the case of the coefficient on x > 0 and the coefficient on $x^2 < 0$, y is increasing in x at first, but will eventually turn around and be decreasing in x (see fig.6.1).
- For the case of the coefficient on x < 0 and the coefficient on $x^2 > 0$, y is decreasing in x at first, but will eventually turn around and be increasing in x (see fig.6.2).
- ♦ For both case, the turning point will be at

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

ometrics

Interaction Terms

- 6-2c Models with Interaction Terms
 - For a model of the form

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

we can't interpret β_1 alone as measuring the change in y with respect to x_1 , we need to take into account β_3 as well, since

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_3 x_2$$

• To summarize the effect of x_1 on y, we typically evaluate the above at average of x_2 .

Econometrics

,

6.3 More of Goodness-of-Fit

Adjusted R-Squared

- Recall that the R² will always increase as more variables are added to the model.
- ◆ The adjusted R² takes into account the number of variables in a model, and may decrease

$$\overline{R}^2 = 1 - \frac{\left[SSR/(n-k-1)\right]}{\left[SST/(n-1)\right]} = 1 - (1-R^2)\frac{n-1}{n-k-1}$$
 (6.22)

Econometrics

Cont. Adjusted R-Squared

- Most PC packages will give you both R^2 and adj- R^2 .
- You can compare the fit of 2 models (with the same y) by comparing the adj- R^2 .
- However, you cannot use the adj- R^2 to compare models with different y's.
 - *e.g. y* vs. ln(*y*).

conometrics

18

R² & Selection of regressors • It is important not to fixate too much on adj R² and lose sight of theory and common sense. • If economic theory clearly predicts a variable belongs, generally leave it in the model. • Don't want to include a variable that prohibits a sensible interpretation of the variable of interest – remember ceteris paribus interpretation of multiple regression. • g. housing price = f (# of rooms, square footage)

Econometrics

