University of Nottingham
UK | CHINA | MALAYSIA

Introductory Econometrics BUSI2053

Further Topics in Regression

(Model specifications)

Lecture Outline

- Modelling Issues (Functional Form)
- Polynomial and Logarithmic Transformations
- Interaction Models (Moderating effect)
- Guides on Choosing a Correct Functional Form

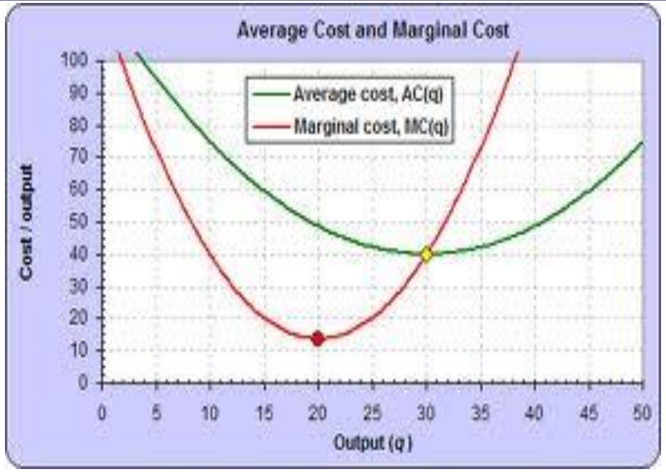
- **Suggested Reading:**
- Chapter 4.3-4.6, 5.6, 5.7, 6.3: Hill, R.C., Griffiths W.E. and Lim, G.C. Principles of Econometrics, fourth edition, Wiley, 2012 (pp. 139-156, pp 189-197)
- Chapter 6.4-6.8, 7.9, 7.10: Gujarati, D.N. and Porter D.C. Basic econometrics, 5th ed., McGraw-Hill, 2009 (pp. 159-173 & pp.207-212)
- Chapter 4: Dougherty, Christopher. Introduction to econometrics. 4th ed. Oxford University Press, 2011 (pp.192-221)
- Chapter 12-13: Westhoff, F. An introduction to econometrics: a self-contained approach, MIT Press, 2013

Modelling Issues (Functional Form)

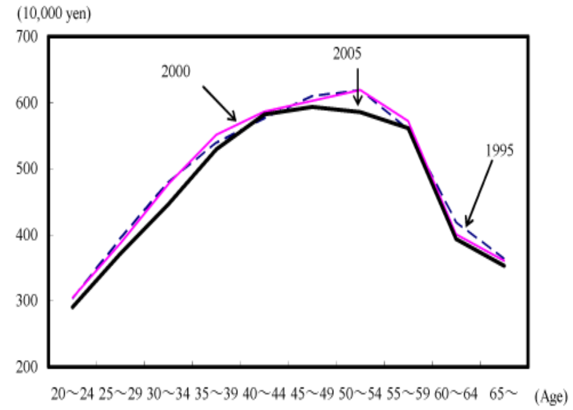
- The first assumption of OLS is linearity.

- According to economic theories, some relationships are not linear

- For example, total cost function, age-wage relationship, marginal revenue function, Cobb-Douglas production function



Appended Figure 2-4 Changes in Age-Wage Curve



Source: *Basic Survey on Wage Structure*, Ministry of Health, Labour and Welfare

Modelling Issues (Functional Form)

- To analyse such non-linear relationships with OLS, we need to transform the variables y and x
- The most common transformations are:
  - Power (polynomial models): If x is a variable, then x^p means raising the variable to the power p (the relationship is not linear in variables)
    - Quadratic (x^2)
    - Cubic (x^3)

  - Natural logarithm: If x is a variable, then its natural logarithm is $\ln(x)$ (the relationship is not linear in coefficients)
    - For example, $Y_i = \beta_1 x_{1i}^{\beta_2} e^{u_i}$

    - Can be transformed as: $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{1i} + u_i$

Polynomial Models

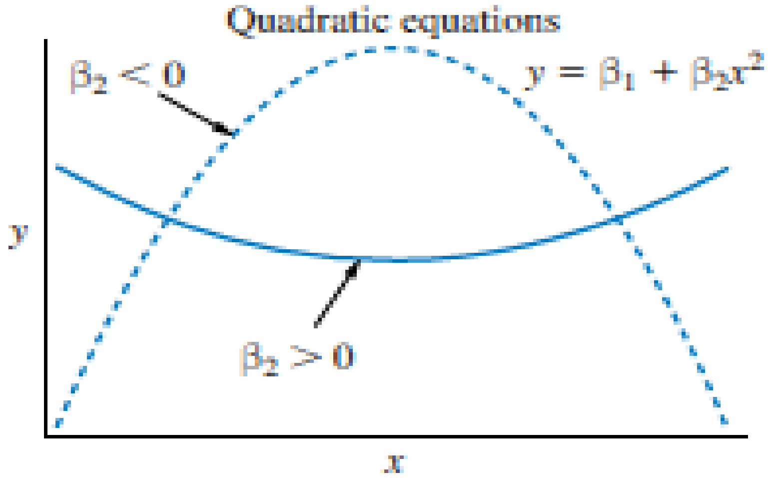
- The general form of a quadratic equation is:

$$y = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + e$$

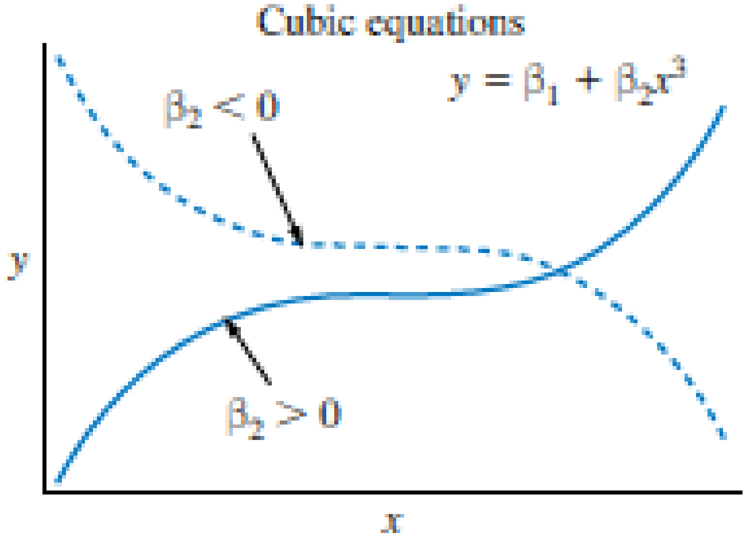
- The general form of a cubic equation is:

$$y = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \beta_4 x_i^3 + e$$

What would happen if the relationship in the above models is not polynomial?



(a)



(b)

Empirical Research

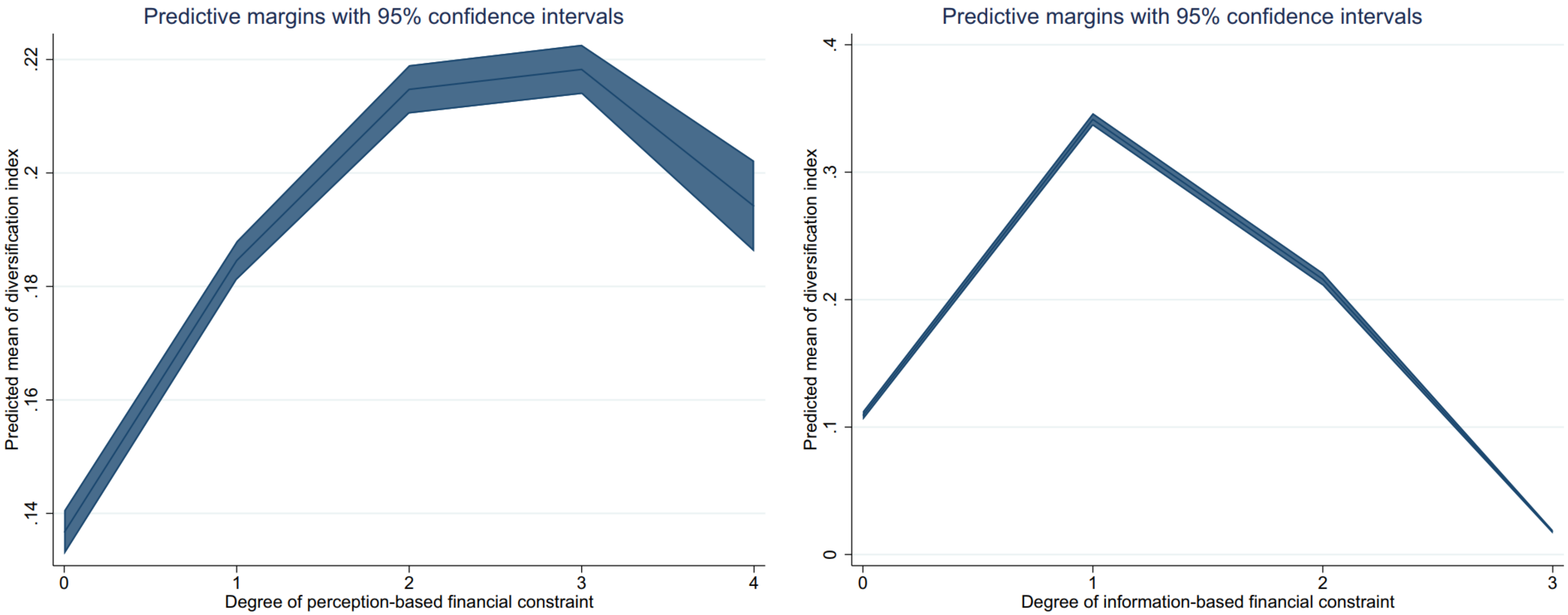


FIGURE 1 The relationship between financial constraint and source diversification. Degree of perception-based financial constraint (0 = No; 1 = Minor; 2 = Moderate; 3 = Major; 4 = Severe). Information-based financial constraint (0 = Not constrained; 1 = Maybe constrained; 2 = Partially constrained; 3 = Fully constrained). The shaded area represents the 95% confidence intervals. [Colour figure can be viewed at wileyonlinelibrary.com]

Polynomial Models (Example)

$$\hat{W} = 3.3248 + 0.4519\,AGE$$

p-value (0.0009) (0.0000)

$$R^2 = 0.0225;\ F(1,\ 7984)=181.72$$

Earning and Age (cps04.gdt)
Stock and Watson, 2nd ed.

Labour theories suggest that wage rate increase with worker's age but at a decreasing rate. So, the relationship should be non-linear. To test the hypothesis of non-linearity, a quadratic term of age is included.

$$\hat{W} = -26.092 + 2.4583\,AGE - 0.0339\,AGE^2$$

p-value (0.0241) (0.0018) (0.0107)

$$R^2 = 0.0231;\ F(2,\ 7983)=94.18$$

What is your conclusion on non-linearity? Why R^2 is very low? At what age does a worker's salary appear to peak?

Polynomial Models (Homework)

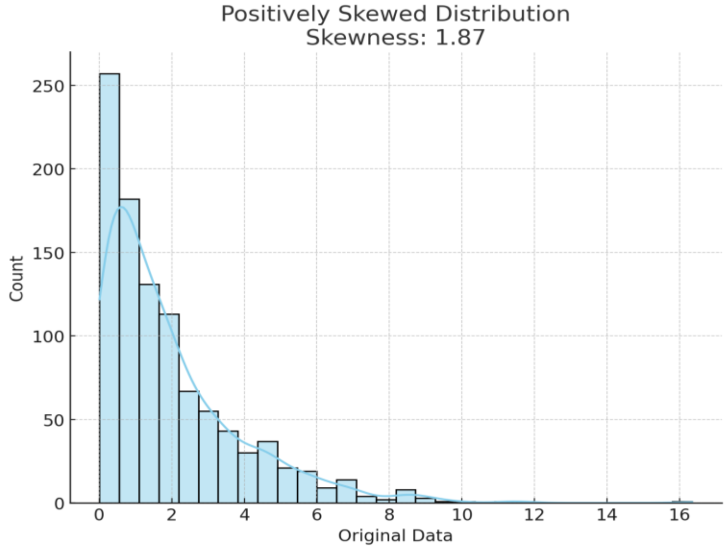
$$\widehat{W} = -26.092 + 2.4583AGE - 0.0339AGE^2$$

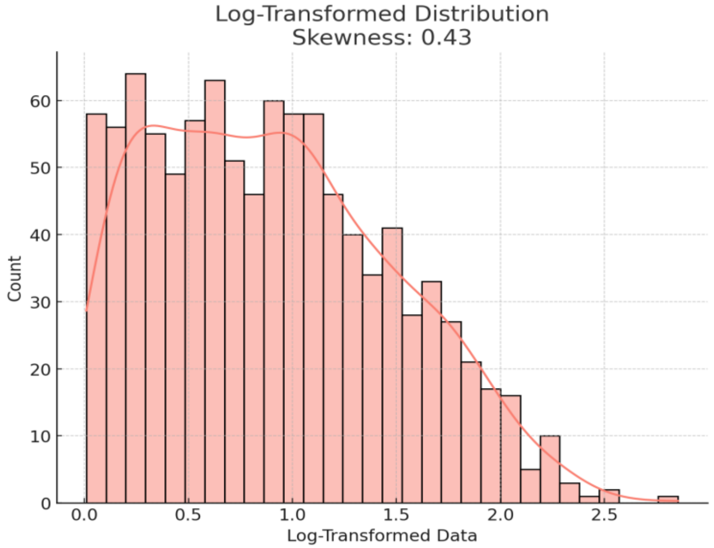
At what age does a worker's salary appear to peak?

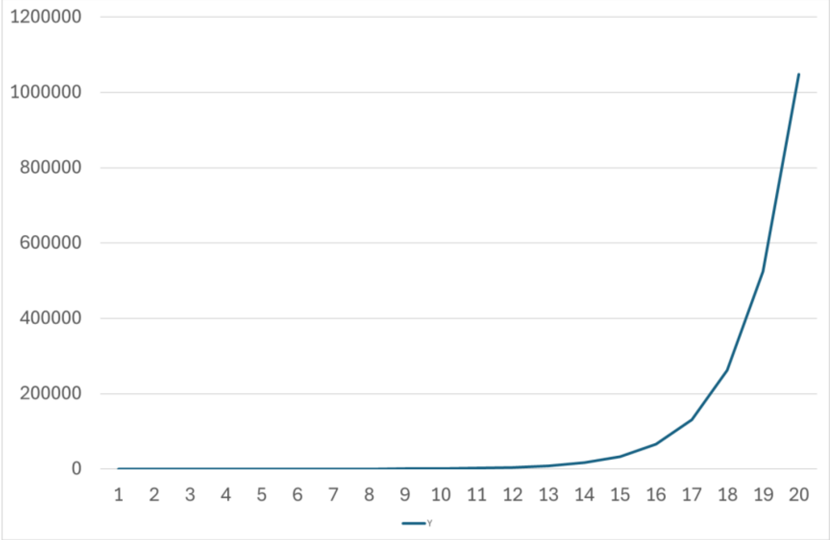
Logarithmic Transformations

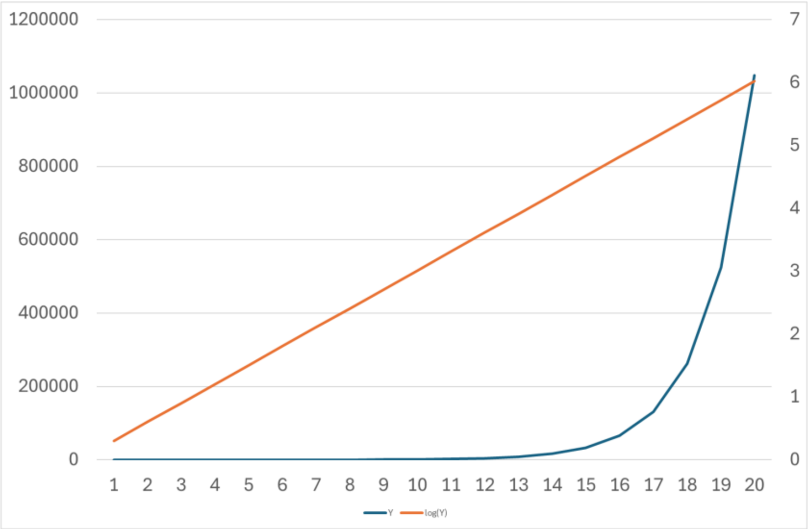
- Econometric models that employ natural logarithms are very common
- Logarithmic transformations are often used for variables that are monetary values
  - Wages, salaries, income, prices, sales, and expenditures
  - In general, for variables that measure the "size" of something
  - These variables have the characteristic that they are positive and often have distributions that are positively skewed, with a long tail to the right
  - Sometimes, **elasticities (% changes)** are more important than unit change in business decision making and the slope coefficient a log-transformed model measures the elasticity Y with respect to X. (Note: change in log represent relative or proportional change)

Logarithmic Transformations









Logarithmic Transformations

- Summary of three configurations:
  1. In the log-log model both the dependent and independent variables are transformed by the "natural" logarithm
     - The parameter β_2 is the **elasticity** of y with respect to x
  2. In the linear-log model the variable x is transformed by the natural logarithm
  3. In the log-linear model only the dependent variable is transformed by the logarithm

Logarithmic Transformations

Interpretation of estimated coefficients

Linear specification: $y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + e$ $$\frac{\partial y}{\partial x_1} = \beta_2$$

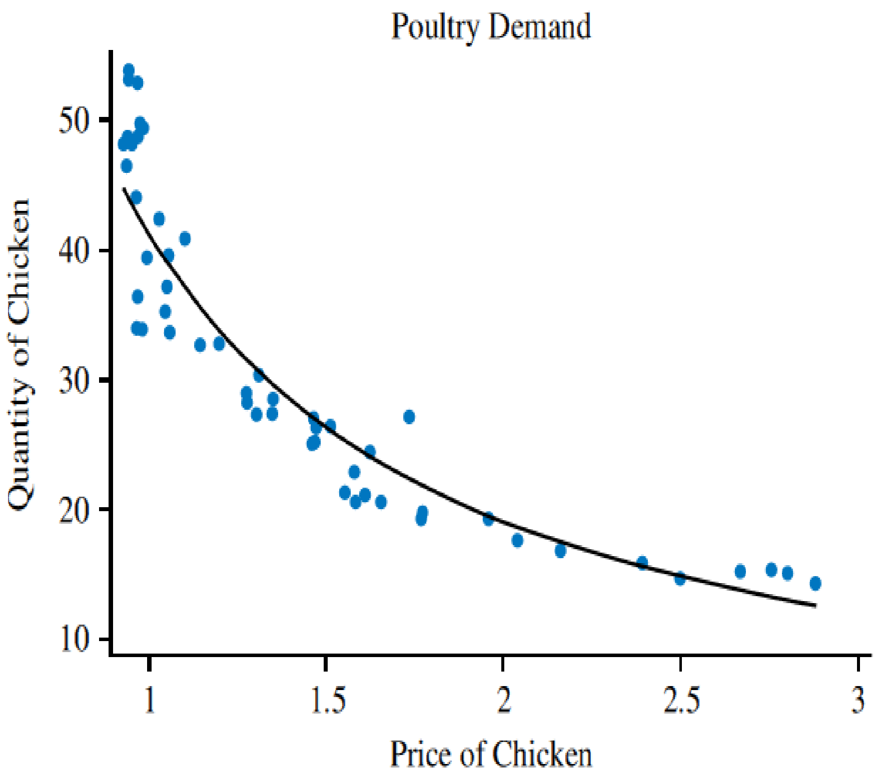
Holding x_2 constant, for every unit change in x_1, y changes by β_2 unit

Log-log specification: $\ln y = \beta_1 + \beta_2 \ln x_1 + \beta_3 \ln x_2 + e$

$$\frac{1}{y}\frac{\partial y}{\partial x_1} = \beta_2 \frac{1}{x_1} \Rightarrow \frac{\partial y}{y} = \beta_2 \frac{\partial x_1}{x} \Rightarrow \beta_2 = \frac{\left(\partial y / y\right)}{\left(\partial x_1 / x_1\right)}$$

Holding x_2 constant, for every 1% change in x_1, y also changes by β_2%
(Also known as constant elasticity model)

Example of a Log-log model



The relationship shows a non-linear shape.

The estimated Log-log model:

$$\widehat{\ln Q} = 3.717 - 1.121 \ln P \qquad R^2 = 0.9136$$
$$\text{s.e.} \quad (0.022) \quad (0.049)$$

We estimate that the price elasticity of demand is 1.121: a 1% increase in real price of fresh chicken is estimated (or expected) to reduce quantity demanded by 1.121%.

Data from *newbroiler.gdt* from POE 4th ed.

Logarithmic Transformations

Interpretation of estimated coefficients

Linear-log specification: $y = \beta_1 + \beta_2 \ln x_1 + \beta_3 \ln x_2 + e$

$$\frac{\partial y}{\partial x_1} = \beta_2 \frac{1}{x_1} \Rightarrow \partial y = \beta_2 \frac{\partial x_1}{x_1} \Rightarrow \beta_2 = \frac{\partial y}{\left({}^{\partial x_1}/_{x_1}\right)}$$

Holding x_2 constant, for every 1% change in x_1, y changes by $0.01 \times \beta_2$ **unit(s)**

Log-linear specification: $\ln y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + e$

$$\frac{1}{y}\frac{\partial y}{\partial x_1} = \beta_2 \Rightarrow \beta_2 = \frac{{}^{\partial y}/_{y}}{\partial x_1}$$

Holding x_2 constant, for every 1 unit change in x_1, y changes by $(100 \times \beta_2)\%$

Example of a Linear-log model

Let's look at our food expenditure model. As Income (x) increase, the coefficient β_2 **is positive** but expected to be decreasing (Why?)

So a linear-log model could be an appropriate functional form: $y = \beta_1 + \beta_2 \ln x_1 + e$

$$\frac{\partial y}{\partial x_1} = \beta_2 \frac{1}{x_1}$$

$$\beta_2 = \frac{\partial y}{\left(\partial x_1 / x_1\right)}$$

```
Model 1: OLS, using observations 1-40
Dependent variable: food_exp

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       -97.1864      84.2374     -1.154    0.2558
  l_income    132.166       28.8046      4.588    4.76e-05 ***

Mean dependent var   283.5735   S.D. dependent var   112.6752
Sum squared resid    318612.4   S.E. of regression   91.56711
R-squared            0.356510   Adjusted R-squared   0.339577
F(1, 38)             21.05302   P-value(F)           0.000048
```

A 1% increase in income will increase expected food expenditure by approximately $1.32 per week.

Example of a Log-linear model

Education and wage (*cps4_small.gdt*)

```
Model 2: OLS, using observations 1-1000
Dependent variable: l_wage

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const       1.60944      0.0864229     18.62    1.15e-066 ***
  educ        0.0904082    0.00614561    14.71    1.75e-044 ***

Mean dependent var   2.856988   S.D. dependent var   0.580619
Sum squared resid    276.7649   S.E. of regression   0.526611
R-squared            0.178205   Adjusted R-squared   0.177381
F(1, 998)            216.4141   P-value(F)           1.75e-44
Log-likelihood      -776.6451   Akaike criterion     1557.290
Schwarz criterion    1567.106   Hannan-Quinn         1561.021

Log-likelihood for wage = -3633.63
```

$$\ln y = \beta_1 + \beta_2 x_1 + e$$

$$\frac{1}{y}\frac{\partial y}{\partial x_1} = \beta_2$$

$$\beta_2 = \frac{\partial y / y}{\partial x_1}$$

An additional year of education increases the expected wage rate by approximately 9%

Notes on Percentage and Percentage point

- Coefficient of a log-log model can be interpreted as percentage change in the dependent variable due to 1 percentage change in the independent variable. Note percentage change in the original variables (not percentage change in logarithm of the variables)

- Example: $\widehat{\ln Q} = 3.717 - 1.121 \ln P$
- Interpretation: A 1% increase in *P* (not ln*P*) is estimated to reduce *Q* (not ln*Q*) by 1.121%.

- If the dependent variable or independent variables are in percentage term, changes can be interpreted as percentage points (not percentage)
- Example: $\ln G\widehat{D}P = 2.717 - 0.00521 Interest\_rate$ where interest rate is in percent
- Interpretation: 1 percentage point increase in interest rate is estimated to reduce *GDP* (not ln*GDP*) by 0.521%.

Interpretations of Log transformations

Model	Example	Interpretation	Elasticity
Log-log	$\ln Y = \beta_1 + \beta_2 \ln X + \varepsilon$	1% change in X will lead to β_2% change in Y.	β_2 (constant)
Log-lin	$\ln Y = \beta_1 + \beta_2 X + \varepsilon$	1 unite change in X will lead to (approximately) $\beta_2 \times 100\%$ change in Y.	$\beta_2 * X$ (not constant)
Lin-log	$Y = \beta_1 + \beta_2 \ln X + \varepsilon$	1% change in X will lead to (approximately) $\beta_2 \times 0.01$ unit change in Y.	β_2 / Y (not constant)
Linear	$Y = \beta_1 + \beta_2 X + \varepsilon$	1 unit change in X will lead to β_2 unit change in Y.	$\beta_2 \times \frac{X}{Y}$ (not constant)

Model specification (Interaction Variable)

- Suppose that we wish to study the effect of income and age on an individual's expenditure on pizza, the initial model would be:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + e$$

The estimated model:

$$\begin{array}{l} PIZZA = 342.88 - 7.576AGE + 1.832INCOME \\ \quad (t) \qquad\qquad\qquad (-3.27) \qquad\quad (3.95) \end{array}$$

The signs of the estimated parameters are as we anticipated Both *AGE* and *INCOME* (in $000) have significant coefficients, based on their *t*-statistics (*pizza4.gdt* from POE 4th ed.)

Model specification (Interaction Variable)

- It is not reasonable to expect that, regardless of the age of the individual, an increase in income by $1,000 should lead to an increase in pizza expenditure by $1.83?

  - It would seem more reasonable to assume that as a person grows older, his or her marginal propensity to spend on pizza declines
    - That is, as a person gets older, less of each extra dollar income is expected to be spent on pizza
  - This is a case in which the effect of income (independent variable) depends on another independent variable, the age of the individual.
    - That is, the effect of one independent variable on the dependent variable is modified by another independent variable.
  - One way of accounting for such interactions is to include an **interaction variable** that is the product of the two independent variables involved

Model specification (Interaction Variables)

- Add the interaction variable (*AGE* x *INCOME*) to the regression model:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4 (AGE \times INCOME) + e$$

- Implications of this revised model are

$$\frac{\partial E(PIZZA)}{\partial AGE} = \beta_2 + \beta_4 INCOME$$

$$\frac{\partial E(PIZZA)}{\partial INCOME} = \beta_3 + \beta_4 AGE$$

- The estimated model is:

$$\widehat{PIZZA} = 161.47 - 2.977AGE + 6.980INCOME - 0.1232(AGE \times INCOME)$$

$$\text{(t)} \qquad (-0.89) \qquad (2.47) \qquad (-1.85)$$

Model specification (Interaction Variables)

$$\widehat{PIZZA} = 161.47 - 2.977AGE + 6.980INCOME - 0.1232(AGE \times INCOME)$$
$$\text{(t)} \qquad\qquad (-0.89) \qquad (2.47) \qquad\qquad (-1.85)$$

- The estimated marginal effect of age upon pizza expenditure for two individuals—one with \$25,000 income and one with \$90,000 income is:

$$\frac{\partial E(PIZZA)}{\partial AGE} = b_2 + b_4 INCOME$$
$$= -2.977 - 0.1232INCOME$$
$$= -6.06 \text{ If } INCOME = 25(\$000)$$
$$= -14.07 \text{ If } INCOME = 90(\$000)$$

  - We expect that an individual with \$25,000 income will reduce pizza expenditures by \$6.06 per year, whereas the individual with \$90,000 income will reduce pizza expenditures by \$14.07 per year

Model specification (Interaction Variables)

Homework

Estimate the marginal effect of income upon pizza expenditure for two individuals—an 18 years old and a 55 years old

$$\widehat{PIZZA} = 161.47 - 2.977AGE + 6.980INCOME - 0.1232(AGE \times INCOME)$$

$$\text{(t)} \qquad (-0.89) \qquad (2.47) \qquad (-1.85)$$

Table 4. Effects of audited financial statement and legal environment on financing source diversification

Variables	Predicted Sign	(1) Fractional Logit	(2) OLS	(3) Fractional Logit	(4) OLS
Audited-Statement	+	0.0224***	0.0227***	0.0223***	0.0221***
		(0.0047)	(0.0047)	(0.0046)	(0.0046)
Legal	+	0.0417**	0.0460**	0.0451***	0.0489***
		(0.0166)	(0.0180)	(0.0165)	(0.0178)
Audited-Statement×*Legal*	–	–		−0.0069***	−0.0064***
				(0.0022)	(0.0021)
Size	+	0.0127***	0.0128***	0.0127***	0.0128***
		(0.0020)	(0.0020)	(0.0020)	(0.0020)
Age	+	0.0331***	0.0282***	0.0324***	0.0276***
		(0.0103)	(0.0094)	(0.0103)	(0.0094)
Age^2	–	−0.0065***	−0.0056***	−0.0063***	−0.0055***
		(0.0020)	(0.0019)	(0.0020)	(0.0019)
Inc	+	−0.0047	−0.0047	−0.0048	−0.0048
		(0.0040)	(0.0040)	(0.0039)	(0.0040)
Bank-Credit	?	−0.0343	−0.0208	−0.0320	−0.0200
		(0.1003)	(0.0904)	(0.0993)	(0.0895)
Constant		–	0.0227***	–	0.0221***
			(0.0047)		(0.0046)
Observations		41,921	41,921	41,921	41,921
R^2		0.0572	0.0562	0.0576	0.0565
Industry dummies		Yes	Yes	Yes	Yes
Country dummies		Yes	Yes	Yes	Yes
Year dummies		Yes	Yes	Yes	Yes

Choosing a Correct Functional Form

1. Choose a shape that is consistent with what economic theory tells us about the relationship.
2. Choose a shape that is sufficiently flexible to "fit" the data.
3. Choose a shape so that assumptions SR1–SR6 are satisfied, ensuring that the least squares estimators have the desirable properties described in Chapters 2 and 3

Summary

After this lecture you should be able to:

- identify appropriate functional form of the model
- interpret the estimated coefficients of polynomial, log-log, log-lin, lin-log and interaction models

University of Nottingham
UK | CHINA | MALAYSIA

Thank you.
Any question?