Nonlinear Relationships

Chapter 7

Prepared by Vera Tabakova, East Carolina University

## Chapter 7: Nonlinear Relationships





## 7.1 Polynomials

### 7.1.1 Cost and Product Curves

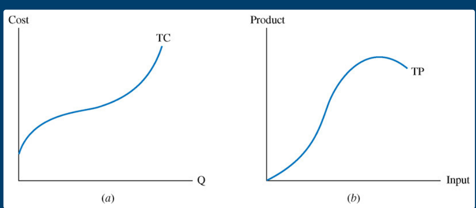


Figure 7.1 (a) Total cost curve and (b) total product curve



## 7.1 Polynomials

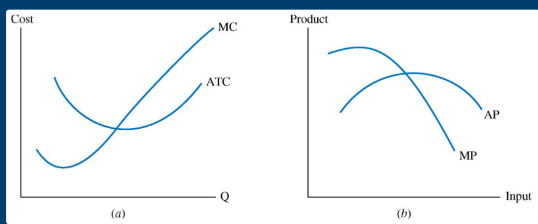


Figure 7.2 Average and marginal (a) cost curves and (b) product curves

## 7.1 Polynomials

$$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + e \quad (7.1)$$

$$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + e \quad (7.2)$$

$$\frac{dE(AC)}{dQ} = \beta_2 + 2\beta_3 Q \quad (7.3)$$

$$\frac{dE(TC)}{dQ} = \alpha_2 + 2\alpha_3 Q + 3\alpha_4 Q^2 \quad (7.4)$$



## 7.1.2 A Wage Equation

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER^2 + e \quad (7.5)$$

$$\frac{\partial E(WAGE)}{\partial EXPER} = \beta_3 + 2\beta_4 EXPER \quad (7.6)$$



## 7.1.2 A Wage Equation

**Table 7.1** Wage Equation with Quadratic Experience

| Variable | Coefficient | Std. Error | $t$-Statistic | Prob. |
|---|---|---|---|---|
| $C$ | −9.8177 | 1.0550 | −9.3062 | 0.0000 |
| $EDUC$ | 1.2101 | 0.0702 | 17.2282 | 0.0000 |
| $EXPER$ | 0.3409 | 0.0514 | 6.6292 | 0.0000 |
| $EXPER^2$ | −0.0051 | 0.0012 | −4.2515 | 0.0000 |

$R^2 = 0.2709$ $SSE = 28420.08$



## 7.1.2 A Wage Equation

$$\left.\frac{\partial E(WAGE)}{\partial EXPER}\right|_{EXPER=18} = .3409 + 2(-.0051)18 = .1576$$

$$EXPER = -\beta_3 / 2\beta_4 = -.3409 / 2(-.0051) = 33.47$$

## 7.2 Dummy Variables

$$PRICE = \beta_1 + \beta_2 SQFT + e \qquad (7.7)$$

$$D = \begin{cases} 1 & \text{if characteristic is present} \\ 0 & \text{if characteristic is not present} \end{cases} \qquad (7.8)$$

$$D = \begin{cases} 1 & \text{if property is in the desirable neighborhood} \\ 0 & \text{if property is not in the desirable neighborhood} \end{cases}$$



## 7.2.1 Intercept Dummy Variables

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + e \qquad (7.9)$$

$$E(PRICE) = \begin{cases} (\beta_1 + \delta) + \beta_2 SQFT & \text{when } D = 1 \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \end{cases} \qquad (7.10)$$



## 7.2.1 Intercept Dummy Variables

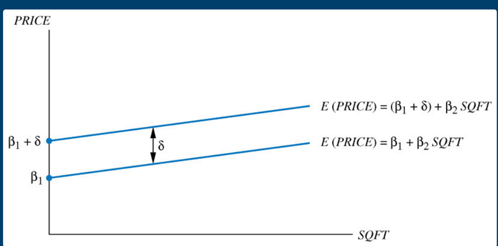


Figure 7.3 An intercept dummy variable



## 7.2.1a Choosing The Reference Group

$$LD = \begin{cases} 1 & \text{if property is not in the desirable neighborhood} \\ 0 & \text{if property is in the desirable neighborhood} \end{cases}$$

$$PRICE = \beta_1 + \lambda LD + \beta_2 SQFT + e$$

$$PRICE = \beta_1 + \delta D + \lambda LD + \beta_2 SQFT + e$$

## 7.2.2 Slope Dummy Variables

$$PRICE = \beta_1 + \beta_2 SQFT + \gamma(SQFT \times D) + e \qquad (7.11)$$

$$E(PRICE) = \beta_1 + \beta_2 SQFT + \gamma(SQFT \times D) = \begin{cases} \beta_1 + (\beta_2 + \gamma)SQFT & \text{when } D = 1 \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \end{cases}$$

$$\frac{\partial E(PRICE)}{\partial SQFT} = \begin{cases} \beta_2 + \gamma & \text{when } D = 1 \\ \beta_2 & \text{when } D = 0 \end{cases}$$



## 7.2.2 Slope Dummy Variables

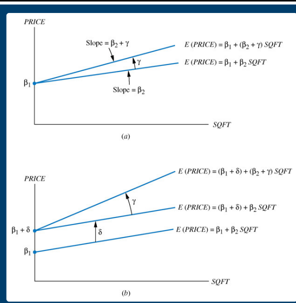


Figure 7.4 (a) A slope dummy variable. (b) A slope and intercept dummy variable



## 7.2.2 Slope Dummy Variables

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + \gamma(SQFT \times D) + e \qquad (7.12)$$

$$E(PRICE) = \begin{cases} (\beta_1 + \delta) + (\beta_2 + \gamma)SQFT & \text{when } D = 1 \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \end{cases}$$



## 7.2.3 An Example: The University Effect on House Prices

**Table 7.2** Representative Real Estate Data Values

| *PRICE* | *SQFT* | *AGE* | *UTOWN* | *POOL* | *FPLACE* |
|---|---|---|---|---|---|
| 205.452 | 23.46 | 6 | 0 | 0 | 1 |
| 185.328 | 20.03 | 5 | 0 | 0 | 1 |
| 248.422 | 27.77 | 6 | 0 | 0 | 0 |
| 287.339 | 23.67 | 28 | 1 | 1 | 0 |
| 255.325 | 21.30 | 0 | 1 | 1 | 1 |
| 301.037 | 29.87 | 6 | 1 | 0 | 1 |

## 7.2.3 An Example: The University Effect on House Prices

$$PRICE = \beta_1 + \delta_1 UTOWN + \beta_2 SQFT + \gamma(SQFT \times UTOWN) + \beta_3 AGE + \delta_2 POOL + \delta_3 FPLACE + e \qquad (7.13)$$



## 7.2.3 An Example: The University Effect on House Prices

*Table 7.3* **House Price Equation Estimates**

| Variable | Coefficient | Std. Error | $t$-Statistic | Prob. |
|---|---|---|---|---|
| $C$ | 24.5000 | 6.1917 | 3.9569 | 0.0001 |
| $UTOWN$ | 27.4530 | 8.4226 | 3.2594 | 0.0012 |
| $SQFT$ | 7.6122 | 0.2452 | 31.0478 | 0.0000 |
| $SQFT \times UTOWN$ | 1.2994 | 0.3320 | 3.9133 | 0.0001 |
| $AGE$ | −0.1901 | 0.0512 | −3.7123 | 0.0002 |
| $POOL$ | 4.3772 | 1.1967 | 3.6577 | 0.0003 |
| $FPLACE$ | 1.6492 | 0.9720 | 1.6968 | 0.0901 |

$R^2 = 0.8706$ $SSE = 230184.4$



## 7.2.3 An Example: The University Effect on House Prices

$$\widehat{PRICE} = (24.5 + 27.453) + (7.6122 + 1.2994)SQFT - .1901AGE + 4.3772POOL + 1.6492FPLACE$$

$$= 51.953 + 8.9116SQFT - .1901AGE + 4.3772POOL + 1.6492FPLACE$$

$$\widehat{PRICE} = 24.5 + 7.6122SQFT - .1901AGE + 4.3772POOL + 1.6492FPLACE$$



## 7.2.3 An Example: The University Effect on House Prices

Based on these regression results, we estimate

- the location premium, for lots near the university, to be \$27,453
- the price per square foot to be \$89.12 for houses near the university, and \$76.12 for houses in other areas.
- that houses depreciate \$190.10 per year
- that a pool increases the value of a home by \$4377.20
- that a fireplace increases the value of a home by \$1649.20

## 7.3 Applying dummy variables

- 7.3.1 Interactions Between Qualitative Factors

$$WAGE = \beta_1 + \beta_2 EDUC + \delta_1 BLACK + \delta_2 FEMALE + \gamma(BLACK \times FEMALE) + e \qquad (7.14)$$

$$E(WAGE) = \begin{cases} \beta_1 + \beta_2 EDUC & WHITE-MALE \\ (\beta_1 + \delta_1) + \beta_2 EDUC & BLACK-MALE \\ (\beta_1 + \delta_2) + \beta_2 EDUC & WHITE-FEMALE \\ (\beta_1 + \delta_1 + \delta_2 + \gamma) + \beta_2 EDUC & BLACK-FEMALE \end{cases}$$



## 7.3.1 Interactions Between Qualitative Factors

**Table 7.4** Wage Equation with Race and Gender

| Variable | Coefficient | Std. Error | *t*-Statistic | Prob. |
|---|---|---|---|---|
| *C* | −3.2303 | 0.9675 | −3.3388 | 0.0009 |
| *EDUC* | 1.1168 | 0.0697 | 16.0200 | 0.0000 |
| *BLACK* | −1.8312 | 0.8957 | −2.0444 | 0.0412 |
| *FEMALE* | −2.5521 | 0.3597 | −7.0953 | 0.0000 |
| *BLACK*×*FEMALE* | 0.5879 | 1.2170 | 0.4831 | 0.6291 |

$R^2 = 0.2482$ $SSE = 29307.71$



## 7.3.1 Interactions Between Qualitative Factors

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

$$\widehat{WAGE} = -4.9122 + 1.1385 EDUC$$
$$\text{(se)} \quad (.9668) \quad (.0716)$$

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(31093 - 29308)/3}{29308/995} = 20.20$$



## 7.3.2 Qualitative Factors with Several Categories

$$WAGE = \beta_1 + \beta_2 EDUC + \delta_1 SOUTH + \delta_2 MIDWEST + \delta_3 WEST + e \qquad (7.15)$$

$$E(WAGE) = \begin{cases} (\beta_1 + \delta_3) + \beta_2 EDUC & WEST \\ (\beta_1 + \delta_2) + \beta_2 EDUC & MIDWEST \\ (\beta_1 + \delta_1) + \beta_2 EDUC & SOUTH \\ \beta_1 + \beta_2 EDUC & NORTHEAST \end{cases}$$

## 7.3.2 Qualitative Factors with Several Categories

**Table 7.5** Wage Equation with Regional Dummy Variables

| Variable | Coefficient | Std. Error | *t*-Statistic | Prob. |
|---|---|---|---|---|
| *C* | −2.4557 | 1.0510 | −2.3365 | 0.0197 |
| *EDUC* | 1.1025 | 0.0700 | 15.7526 | 0.0000 |
| *BLACK* | −1.6077 | 0.9034 | −1.7795 | 0.0755 |
| *FEMALE* | −2.5009 | 0.3600 | −6.9475 | 0.0000 |
| *BLACK×FEMALE* | 0.6465 | 1.2152 | 0.5320 | 0.5949 |
| *SOUTH* | −1.2443 | 0.4794 | −2.5953 | 0.0096 |
| *MIDWEST* | −0.4996 | 0.5056 | −0.9880 | 0.3234 |
| *WEST* | −0.5462 | 0.5154 | −1.0597 | 0.2895 |

$R^2 = 0.2535$ $SSE = 29101.3$



## 7.3.3 Testing the Equivalence of Two Regressions

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + \gamma(SQFT \times D) + e$$

$$E(PRICE) = \begin{cases} \alpha_1 + \alpha_2 SQFT & D = 1 \\ \beta_1 + \beta_2 SQFT & D = 0 \end{cases}$$

$$WAGE = \beta_1 + \beta_2 EDUC + \delta_1 BLACK + \delta_2 FEMALE + \gamma(BLACK \times FEMALE) + e$$



## 7.3.3 Testing the Equivalence of Two Regressions

$$\begin{aligned} WAGE = {} & \beta_1 + \beta_2 EDUC + \delta_1 BLACK + \delta_2 FEMALE + \gamma(BLACK \times FEMALE) + \\ & \theta_1 SOUTH + \theta_2(EDUC \times SOUTH) + \theta_3(BLACK \times SOUTH) + \\ & \theta_4(FEMALE \times SOUTH) + \theta_5(BLACK \times FEMALE \times SOUTH) + e \end{aligned} \quad (7.16)$$

$$E(WAGE) = \begin{cases} \beta_1 + \beta_2 EDUC + \delta_1 BLACK + \delta_2 FEMALE + \\ \gamma(BLACK \times FEMALE) & SOUTH = 0 \\ (\beta_1 + \theta_1) + (\beta_2 + \theta_2) EDUC + (\delta_1 + \theta_3) BLACK + \\ (\delta_2 + \theta_4) FEMALE + (\gamma + \theta_5)(BLACK \times FEMALE) & SOUTH = 1 \end{cases}$$



## 7.3.3 Testing the Equivalence of Two Regressions

**Table 7.6** Comparison of Fully Interacted to Separate Models

| | (1) Full sample | | (2) Non-south | | (3) South | |
|---|---|---|---|---|---|---|
| Variable | Coefficient | Std. Error | Coefficient | Std. Error | Coefficient | Std. Error |
| *C* | −3.5775 | 1.1513 | −3.5775 | 1.2106 | −2.2752 | 1.5550 |
| *EDUC* | 1.1658 | 0.0824 | 1.1658 | 0.0866 | 0.9741 | 0.1143 |
| *BLACK* | −0.4312 | 1.3482 | −0.4312 | 1.4176 | −2.1756 | 1.0804 |
| *FEMALE* | −2.7540 | 0.4257 | −2.7540 | 0.4476 | −1.8421 | 0.5896 |
| *BLACK×FEMALE* | 0.0673 | 1.9063 | 0.0673 | 2.0044 | 0.6101 | 1.4329 |
| *SOUTH* | 1.3023 | 2.1147 | | | | |
| *EDUC×SOUTH* | −0.1917 | 0.1542 | | | | |
| *BLACK×SOUTH* | −1.7444 | 1.8267 | | | | |
| *FEMALE×SOUTH* | 0.9119 | 0.7960 | | | | |
| *BLACK×FEMALE×SOUTH* | 0.5428 | 2.5112 | | | | |
| *SSE* | 29012.7 | | 22031.3 | | 6981.4 | |
| *N* | 1000 | | 685 | | 315 | |

## 7.3.3 Testing the Equivalence of Two Regressions

$$H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$$

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(29307.7 - 29012.7)/5}{29012.7/990} = 2.0132$$



## 7.3.3 Testing the Equivalence of Two Regressions

**Remark:** The usual *F*-test of a joint hypothesis relies on the assumptions MR1-MR6 of the linear regression model. Of particular relevance for testing the equivalence of two regressions is assumption MR3, that the variance of the error term is the same for all observations. If we are considering possibly different slopes and intercepts for parts of the data, it might also be true that the error variances are different in the two parts of the data. In such a case the usual *F*-test is not valid.



## 7.3.4 Controlling for Time

*7.3.4a Seasonal Dummies*

*7.3.4b Annual Dummies*

*7.3.4c Regime Effects*

$$ITC = \begin{cases} 1 & 1962-1965, 1970-1986 \\ 0 & otherwise \end{cases}$$

$$INV_t = \beta_1 + \delta ITC_t + \beta_2 GNP_t + \beta_3 GNP_{t-1} + e_t$$



## 7.4 Interactions Between Continuous Variables

*Table 7.7* **Pizza Expenditure Data**

| *PIZZA* | *INCOME* | *AGE* |
|---|---|---|
| 109 | 15000 | 25 |
| 0 | 30000 | 45 |
| 0 | 12000 | 20 |
| 108 | 20000 | 28 |
| 220 | 15000 | 25 |

## 7.4 Interactions Between Continuous Variables

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + e \quad (7.17)$$

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

$$\partial E(PIZZA)/\partial INCOME = \beta_3$$

$$\widehat{PIZZA} = 342.88 - 7.58 AGE + .0024 INCOME$$
$$(t) \qquad (-3.27) \qquad (3.95)$$



## 7.4 Interactions Between Continuous Variables

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4 (AGE \times INCOME) + e \quad (7.18)$$

$$\partial E(PIZZA)/\partial AGE = \beta_2 + \beta_4 INCOME$$

$$\partial E(PIZZA)/\partial INCOME = \beta_3 + \beta_4 AGE$$

$$\widehat{PIZZA} = 161.47 - 2.98 AGE + .009 INCOME - .00016 (AGE \times INCOME)$$
$$(t) \qquad (-.89) \qquad (2.47) \qquad (-1.85)$$



## 7.4 Interactions Between Continuous Variables

$$\frac{\widehat{\partial E(PIZZA)}}{\partial AGE} = b_2 + b_4 INCOME$$

$$= -2.98 - .00016 INCOME$$

$$= \begin{cases} -6.98 & \text{for } INCOME = \$25{,}000 \\ -17.40 & \text{for } INCOME = \$90{,}000 \end{cases}$$



## 7.5 Log-Linear models

7.5.1 Dummy Variables

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \delta FEMALE \quad (7.19)$$

$$\ln(WAGE) = \begin{cases} \beta_1 + \beta_2 EDUC & MALES \\ (\beta_1 + \delta) + \beta_2 EDUC & FEMALES \end{cases}$$

## 7.5.1a A Rough Calculation

$$\ln(WAGE)_{FEMALES} - \ln(WAGE)_{MALES} = \Delta \ln\left(WAGE\right) = \delta$$

$$\begin{aligned}\widehat{\ln(WAGE)} &= .9290 + .1026EDUC - .2526FEMALE \\ \text{(se)} \quad & \quad\ (.0837) \quad (.0061) \qquad\quad (.0300)\end{aligned}$$



## 7.5.1b An Exact Calculation

$$\ln(WAGE)_{FEMALES} - \ln(WAGE)_{MALES} = \ln\left(\frac{WAGE_{FEMALES}}{WAGE_{MALES}}\right) = \delta$$

$$\frac{WAGE_{FEMALES}}{WAGE_{MALES}} = e^{\delta}$$

$$\frac{WAGE_{FEMALES}}{WAGE_{MALES}} - \frac{WAGE_{MALES}}{WAGE_{MALES}} = \frac{WAGE_{FEMALES} - WAGE_{MALES}}{WAGE_{MALES}} = e^{\delta} - 1$$

$$100\left(e^{\delta} - 1\right)\% = 100\left(e^{-.2526} - 1\right)\% = -22.32\%$$



## 7.5.2 Interaction and Quadratic Terms

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \gamma(EDUC \times EXPER) \qquad (7.20)$$

$$\left.\frac{\Delta \ln(WAGE)}{\Delta EXPER}\right|_{EDUC\ fixed} = \beta_3 + \gamma EDUC$$

$$\begin{aligned}\widehat{\ln(WAGE)} &= .1528 + .1341EDUC + .0249EXPER - .000962(EDUC \times EXPER) \\ \text{(se)} \quad & \quad (.1722)\ (.0127) \qquad\quad (.0071) \qquad\quad (.00054)\end{aligned}$$



## 7.5.2 Interaction and Quadratic Terms

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER^2 + \gamma\left(EDUC \times EXPER\right)$$

$$\%\Delta WAGE \cong 100\left(\beta_3 + 2\beta_4 EXPER + \gamma EDUC\right)\%$$

## Keywords

- annual dummy variables
- binary variable
- Chow test
- collinearity
- dichotomous variable
- dummy variable
- dummy variable trap
- exact collinearity
- hedonic model
- interaction variable
- intercept dummy variable
- log-linear models
- nonlinear relationship
- polynomial
- reference group
- regional dummy variable
- seasonal dummy variables
- slope dummy variable



## Chapter 7 Appendix

- Appendix 7 Details of log-linear model interpretation



## Appendix 7
## Details of log-linear model interpretation

$$E(y) = \exp(\beta_1 + \beta_2 x + \sigma^2/2) = \exp(\beta_1 + \beta_2 x) \times \exp\left(\sigma^2/2\right)$$

$$E(y) = \exp(\beta_1 + \beta_2 x + \delta D) \times \exp\left(\sigma^2/2\right)$$



## Appendix 7
## Details of log-linear model interpretation

$$\%\Delta E(y) = 100\left[\frac{E(y_1) - E(y_0)}{E(y_0)}\right]\%$$

$$= 100\left[\frac{\left(\exp(\beta_1 + \beta_2 x + \delta) \times \exp\left(\sigma^2/2\right)\right) - \left(\exp(\beta_1 + \beta_2 x) \times \exp\left(\sigma^2/2\right)\right)}{\left(\exp(\beta_1 + \beta_2 x) \times \exp\left(\sigma^2/2\right)\right)}\right]\%$$

$$= 100\left[\frac{\exp(\beta_1 + \beta_2 x)\exp(\delta) - \exp(\beta_1 + \beta_2 x)}{\exp(\beta_1 + \beta_2 x)}\right]\%$$

$$= 100\left[\exp(\delta) - 1\right]\%$$

## Appendix 7
## Details of log-linear model interpretation

$$E(y) = \exp(\beta_1 + \beta_2 x + \beta_3 z + \gamma(xz)) \times \exp(\sigma^2/2)$$

$$\frac{\partial E(y)}{\partial z} = \exp(\beta_1 + \beta_2 x + \beta_3 z + \gamma(xz)) \times \exp(\sigma^2/2) \times (\beta_3 + \gamma x))$$

$$100\left[\frac{\partial E(y)/E(y)}{\partial z}\right] = 100(\beta_3 + \gamma x)\%$$