Chapter 3 Functional Form, Nonlinearity and Specification

(5th Chap.7 、Chap.8 and Chap.9)

3.1 Dummy Variables

3.1.1 Comparing Two means

two groups:

$$\begin{cases} Group1\ y_i = \mu + \varepsilon_i \Rightarrow \bar{y}_1 = \hat{\mu} \text{ since } y_i = \hat{\mu} + e_i \text{ with } \sum e_i = 0 \\ Group2\ y_i = \mu + \delta + \varepsilon_i \Rightarrow \bar{y}_2 = \hat{\mu} + \hat{\delta} \Rightarrow \hat{\delta} = \bar{y}_2 - \hat{\mu} = \bar{y}_2 - \bar{y}_1 \end{cases}$$

Combined in a single equation by a dummy variable:

1). One dummy

approach I:

$$d_i = \begin{cases} 0 \text{ if group } 1 \to n_2 \\ 1 \text{ if group } 2 \to n_1 \end{cases}$$

$$\Rightarrow y_i = \mu + d_i\delta + \varepsilon_i, \quad \text{where } X = (i, d) \qquad (3.1)$$

$$(X'X) = \begin{pmatrix} i' \\ d' \end{pmatrix}(i, d) = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}$$

$$\Rightarrow (X'X)^{-1} = \frac{1}{nn_1 - n_1^2}\begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix} = \frac{1}{n_1 n_2}\begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix}$$

$$X'Y = \begin{pmatrix} i'Y \\ d'Y \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum d_i y_i \end{pmatrix}$$

$$\Rightarrow (X'X)^{-1} X'Y = \frac{1}{n_1 n_2}\begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix}\begin{pmatrix} \sum y_i \\ \sum d_i y_i \end{pmatrix} = \frac{1}{n_1 n_2}\begin{pmatrix} n_1 \sum (1 - d_i) y_i \\ n \sum d_i y_i - n_1 \sum y_i \end{pmatrix}$$

$$\begin{pmatrix} \hat{\mu} \\ \hat{\delta} \end{pmatrix} = (X'X)^{-1} X'Y = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \end{pmatrix}$$

2).Two dummies and dummy trap

$$d_i = \begin{cases} 0 \text{ if group } 1 \\ 1 \text{ if group } 2 \end{cases} \qquad h_i = \begin{cases} 1 \text{ if group } 1 \\ 0 \text{ if group } 2 \end{cases} = 1 - d_i$$

Model $y_i = \mu_1 + \mu_2 d_i + \mu_3 h_i + \varepsilon_i$ (3.2)

The two dummy variables d_i and h_i sum to one at every observation, which would reproduce the constant term—a case of perfect multicollinearity. This is known as the **dummy variable trap**.

Conclusion: if there are n groups, you should only use n-1 dummies. Or, you could use n dummies while without intercept.

approach II:

$$y_i = \mu_1 h_i + \mu_2 d_i + \varepsilon_i \qquad (3.3)$$

$$X = \begin{pmatrix} i_1 & 0 \\ 0 & i_2 \end{pmatrix}$$

$$\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = (X'X)^{-1} X'Y = \left[\begin{pmatrix} i_1' & 0 \\ 0 & i_2' \end{pmatrix} \begin{pmatrix} i_1 & 0 \\ 0 & i_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} i_1' & 0 \\ 0 & i_2' \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n_1} y_{1i} \\ \sum_{i=1}^{n_2} y_{2i} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$$

Which approach is more appropriate depends on purpose. In general, approach I is more appropriate in most cases. While in particular, especially in panel data model, approach II may be more appropriate.

3.1.2 Different types of Dummy Variables

1) Different intercepts, common slopes

$$Y_i = \beta_1 + \beta_2 X_{2i} + \alpha d_i + \varepsilon_i \Rightarrow \begin{cases} Y_i = \beta_1 + \beta_2 X_{2i} + \varepsilon_i \\ Y_i = (\beta_1 + \alpha) + \beta_2 X_{2i} + \varepsilon_i \end{cases} \qquad (3.4)$$

$\Leftrightarrow$ Equivalent to test: $\alpha = 0$ by Chow Test.

2) Different slopes, common intercepts

$$Y_i = \beta_1 + \beta_2 X_{2i} + \alpha d_i X_{2i} + \varepsilon_i \Rightarrow \begin{cases} Y_i = \beta_1 + \beta_2 X_{2i} + \varepsilon_i \\ Y_i = \beta_1 + (\beta_2 + \alpha) X_{2i} + \varepsilon_i \end{cases} \qquad (3.5)$$

$\Leftrightarrow$ Equivalent to test: $\alpha = 0$ by Chow Test

3) Different intercepts, different slopes

$$Y_i = \beta_1 + \beta_2 X_{2i} + \alpha_1 d_i + \alpha_2 d_i X_{2i} + \varepsilon_i \Rightarrow \begin{cases} Y_i = \beta_1 + \beta_2 X_{2i} + \varepsilon_i \\ Y_i = (\beta_1 + \alpha_1) + (\beta_2 + \alpha_2) X_{2i} + \varepsilon_i \end{cases} \quad (3.6)$$

$\Leftrightarrow$ Equivalent to test: $\alpha_1 = \alpha_2 = 0$ by Chow Test.

3.1.3 Quarterly data

$$d_1 = \begin{cases} 1 & \text{Quarter 1} \\ 0 & \text{Other} \end{cases} \quad d_2 = \begin{cases} 1 & \text{Quarter 2} \\ 0 & \text{Other} \end{cases} \quad d_3 = \begin{cases} 1 & \text{Quarter 3} \\ 0 & \text{Other} \end{cases} \quad d_4 = \begin{cases} 1 & \text{Quarter 4} \\ 0 & \text{Other} \end{cases}$$

Model I: $Y = \beta_1 i + \beta_2 d_2 + \beta_3 d_3 + \beta_4 d_4 + X\alpha + \varepsilon$ (3.7)

Model II: $Y = \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 d_4 + X\alpha + \varepsilon$ (3.8)

$\hat{\beta}_1$ and $\hat{\alpha}$ are the same for model I and model II, while $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ are different.

Seasonal effects $X_1 = (d_1 \quad d_2 \quad d_3 \quad d_4)$, $X_2 = X$

Model III: $Y^* = \beta_5 X_2^* + \varepsilon$ (3.9)

Where $X_2^* = M_1 X_2$, $Y^* = M_1 Y$, $M_1 = I - X_1 (X_1' X_1)^{-1} X_1'$

Deseasonalizing: $\hat{\beta}_5 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y = \left(X_2^{*\prime} X_2^*\right)^{-1} X_2^* Y^*$

Neither Y^* and X_2^* has seasonal effects.

3.1.4 Threshold effects

$$Income = \beta_1 + \beta_2 \cdot age + \beta_3 \cdot Edu + \varepsilon$$

$$Edu = \begin{cases} 1 & \text{High School} \\ 2 & \text{Bachelor} \\ 3 & \text{Master} \\ 4 & \text{Ph.D.} \end{cases}$$

This approach is unreasonable, because it assumes that the marginal effects of income are the same for each level of education. There are two methods to solve this problem, see 5^{th} ,Greene, P120.

3.1.5. Spline Regression

$$Y = \beta_1 + \beta_2 X_{(X < t_1^*)} + \delta_1 X_{(t_1^* \le X < t_2^*)} + \delta_2 X_{(X \ge t_2^*)} + \varepsilon \quad (3.10)$$

Step 1: using dummy variables to denote thresholds:

$$d_1 = \begin{cases} 1 \quad if \ \ age \geq 18 \ldots \left(t_1^*\right) \\ 0 \quad if \ \ age < 18 \ldots \left(t_1^*\right) \end{cases} \tag{3.11}$$

$$d_2 = \begin{cases} 1 \quad if \ \ age \geq 22 \ldots \left(t_2^*\right) \\ 0 \quad if \ \ age < 22 .. \left(t_2^*\right) \end{cases} \tag{3.12}$$

Basic model to illustrate threshold effects

$$Income = \beta_1 + \beta_2 age + \gamma_1 d_1 + \delta_1 d_1 age + \gamma_2 d_2 + \delta_2 d_2 age + \varepsilon$$

Step 2: add continuity restrictions:

$$\begin{cases} \beta_1 + \beta_2 t_1^* = \left(\beta_1 + \gamma_1\right) + \left(\beta_2 + \delta_1\right) t_1^* \\ \left(\beta_1 + \gamma_1\right) + \left(\beta_2 + \delta_1\right) t_2^* = \left(\beta_1 + \gamma_1 + \gamma_2\right) + \left(\beta_2 + \delta_1 + \delta_2\right) t_2^* \end{cases} \Rightarrow \gamma_1 = -\delta_1 t_1^*, \gamma_2 = -\delta_2 t_2^*$$

The Threshold value, 18 and 22 are called knots. Then

$$\begin{aligned} E\left(Income \middle| age\right) &= \beta_1 + \beta_2 age \quad if \ \ age < 18 \\ &= \beta_1 + \beta_2 age + \delta_1 \left(age - 18\right) \quad if \ \ 18 \leq age < 22 \\ &= \beta_1 + \beta_2 age + \delta_1 \left(age - 18\right) + \delta_2 \left(age - 22\right) \quad if \ \ age \geq 22 \end{aligned}$$

when $age = 18,\ E\left(Income\right) = \beta_1 + \beta_2 age$;

when $age = 22,\ E\left(income\right) = \beta_1 + \beta_2 age + \delta_1 \left(age - 18\right)$

3.1.6 Interaction terms

e.g. $Y = \beta_1 + \beta_2 d_2 + \beta_3 X_3 + \varepsilon$

$$d_2 = \begin{cases} 1 \ \ \text{export promoting} \\ 0 \ \ \text{import substituting} \end{cases}, \quad X_3 = FDI$$

$\frac{\partial Y}{\partial X_3} = \beta_3$, which means marginal effect of X_3 on Y will be the same for the two groups, it is not real. That’s why we introduce “interaction terms”.

$$Y = \beta_1 + \beta_2 d_2 + \beta_3 X_3 + \beta_4 d_2 X_3 + \varepsilon$$

$$\Rightarrow \frac{\partial Y}{\partial X_3} = \beta_3 + \beta_4 d_2 = \begin{cases} \beta_3 + \beta_4 \ \ if \ \ d_2 = 1 \\ \beta_3 \ \ if \ \ d_2 = 0 \end{cases} \rightarrow different$$

Cause to introduce “interaction terms” is that the marginal effect is different when other variables are at different levels. i.e. depends on other variables.

3.2 Specification Analysis

Our analysis has been based on the assumption that the correct specification of the regression model is known to be: $Y = X\beta + \varepsilon$

3.2.1 Omission of Relevant Variables

The true model: $Y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$ （3.12）

The specified model: $Y = X_1\beta_1 + \varepsilon$ （3.13）

$$\hat{\beta}_1 = \left(X_1^{'}X_1\right)^{-1} X_1^{'}Y$$

$$= \left(X_1^{'}X_1\right)^{-1} X_1^{\prime}\left(X_1\beta_1 + X_2\beta_2 + \varepsilon\right)$$

$$= \beta_1 + \left(X_1^{'}X_1\right)^{-1} X_1^{'}X_2\beta_2 + \left(X_1^{'}X_1\right)^{-1} X_1^{'}\varepsilon$$

1) $E\left(\hat{\beta}_1 \middle| X\right) = \beta_1 + \left(X_1^{'}X_1\right)^{-1} X_1^{'}X_2\beta_2$ （3.14）

Unless $X_1^{'}X_2 = 0$, $\hat{\beta}_1$ is biased.

Example:

$$\left.\begin{array}{l} y_i = \beta_1 + x_{2i}\beta_2 + \varepsilon_i \Rightarrow y_i - \overline{y} = \left(x_{2i} - \overline{x}_2\right)\beta_2 + \varepsilon_i^* \Rightarrow \hat{\beta}_2 = \dfrac{\sum\left(x_{2i} - \overline{x}_2\right)\left(y_i - \overline{y}\right)}{\sum\left(x_{2i} - \overline{x}_2\right)^2} \\ y_i = \beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \varepsilon_i \Rightarrow y_i - \overline{y} = \left(x_{2i} - \overline{x}_2\right)\beta_2 + \left(x_{3i} - \overline{x}_3\right)\beta_3 + \varepsilon_i^* \end{array}\right\}$$

$$\Rightarrow \hat{\beta}_2 = \beta_2 + \frac{\sum\left(x_{2i} - \overline{x}_2\right)\left(x_{3i} - \overline{x}_3\right)}{\sum\left(x_{2i} - \overline{x}_2\right)^2}\beta_3 + \frac{\sum\left(x_{2i} - \overline{x}_2\right)\varepsilon_i^*}{\sum\left(x_{2i} - \overline{x}_2\right)^2}$$

$$\Rightarrow E\left(\hat{\beta}_2 \middle| X\right) = \beta_2 + \frac{\sum\left(x_{2i} - \overline{x}_2\right)\left(x_{3i} - \overline{x}_3\right)}{\sum\left(x_{2i} - \overline{x}_2\right)^2}\beta_3$$

$$= \beta_2 + \frac{Cov\left(x_2, x_3\right)}{Var\left(x_2\right)}\beta_3$$

① If $Cov\left(x_2, x_3\right) > 0$ and $\beta_3 > 0$, or $Cov\left(x_2, x_3\right) < 0$ and $\beta_3 < 0$ then $E\left(\hat{\beta}_2 \middle| X\right) > \beta_2$, i.e. $\hat{\beta}_2$ is upper-biased.

② If $Cov\left(x_2, x_3\right) > 0$ and $\beta_3 < 0$, or $Cov\left(x_2, x_3\right) < 0$ and $\beta_3 > 0$, then $E\left(\hat{\beta}_2 \middle| X\right) < \beta_2$, i.e. $\hat{\beta}_2$ is lower-biased.

2) $e_1 = M_1 Y = M_1\left(X_1\beta_1 + X_2\beta_2 + \varepsilon\right) = M_1 X_2\beta_2 + M_1\varepsilon$

$$e_1'e_1 = \beta_2' X_2' M_1 X_2 \beta_2 + \varepsilon' M_1 \varepsilon + \varepsilon' M_1 X_2 \beta_2 + \beta_2' X_2' M_1 \varepsilon$$

$$E\left(e_1'e_1 \middle| X\right) = \left(n - K_1\right)\sigma^2 + \beta_2' X_2' M_1 X_2 \beta_2$$

$$E\left(\frac{e_1'e_1}{n-K_1}\middle| X\right) = \sigma^2 + \frac{\beta_2' X_2' M_1 X_2 \beta_2}{n-K_1} \geq \sigma^2 \qquad (3.15)$$

s^2 is biased upward.

3) $Var\left(\hat{\beta}_1 \middle| X\right) = \sigma^2\left(X_1'X_1\right)^{-1}$

$$Var\left(\hat{\beta}_{1.2} \middle| X\right) = \sigma^2\left(X_1' M_2 X_1\right)^{-1}$$

$$\because Var\left(\hat{\beta}_{1.2} \middle| X\right) = \sigma^2\left(\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}\left(X_1, X_2\right)\right)^{-1}_{upper.left.submatrix}$$

$$= \sigma^2\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1}_{upper.left.submatrix}$$

$$= \sigma^2\left[X_1'X_1 - X_1'X_2\left(X_2'X_2\right)^{-1}X_2'X_1\right]^{-1}$$

$$= \sigma^2\left(X_1'M_2X_1\right)^{-1} \quad \left(M_2 = I - X_2\left(X_2'X_2\right)^{-1}X_2'\right)$$

$$Var\left(\hat{\beta}_1 \middle| X\right)^{-1} - Var\left(\hat{\beta}_{1.2} \middle| X\right)^{-1} = \left(1/\sigma^2\right)X_1'X_2\left(X_2'X_2\right)^{-1}X_2'X_1 \qquad (3.16)$$

which is nonnegative definite.

Although $\hat{\beta}_1$ is biased, its variance is never lager than that of $\hat{\beta}_{1.2}$.

For example, X_1 and X_2 are each a single column and that the variables are measured as deviations from their respective means. Then

$Var\left(\hat{\beta}_1 \middle| X\right) = \dfrac{\sigma^2}{s_{11}}$, where $s_{11} = \sum\left(X_{i1} - \overline{X}_1\right)^2 = \sum x_{i1}^2 = x_1'x_1$,

$Var\left(\hat{\beta}_{1.2} \middle| X\right) = \sigma^2\left[x_1'x_1 - x_1'x_2\left(x_2'x_2\right)^{-1}x_2'x_1\right]^{-1} = \dfrac{\sigma^2}{s_{11}\left(1 - r_{12}^2\right)}$, where $r_{12}^2 = \dfrac{\left(x_1'x_2\right)^2}{x_1'x_1x_2'x_2}$

The more highly correlated x_1 and x_2 are, the larger is the variance of $\hat{\beta}_{1.2}$ compared with that of $\hat{\beta}_1$. So, it is possible that $\hat{\beta}_1$ is a more precise estimator based on the mean-squared-error criterion.

4) Conclusion:

When we omit relevant variables in regression, the estimators of β_1 and σ^2 will be biased. Since we can not estimate β_1 and σ^2 unbiasedly, we can not test the hypothesis of β_1 correctly.

3.2.2 Inclusion of Irrelevant Variables

The true model: $Y = X_1\beta_1 + \varepsilon$ （3.17）

The specified model: $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ （3.18）

(X_2 : irrelevant variables)

1) $\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X_1\beta + \varepsilon)$

$$E\left(\hat{\beta}\middle|X\right) = E\left(\begin{matrix}\hat{\beta}_1\\ \hat{\beta}_2\end{matrix}\middle|X\right) = (X'X)^{-1}X'X_1\beta_1 = (X'X)^{-1}X'X\begin{pmatrix}I_{K_1}\\ 0\end{pmatrix}\beta_1 = \begin{pmatrix}\beta_1\\ 0\end{pmatrix}$$

$\therefore E\left(\hat{\beta}_1\middle|X\right) = \beta_1,\ E\left(\hat{\beta}_2\middle|X\right) = 0 = \beta_2$ （3.19）

2) $Var\left(\hat{\beta}_1\middle|X\right) = \sigma^2(X_1'X_1)^{-1}$

$Var\left(\hat{\beta}_{1,2}\middle|X\right) = \sigma^2(X_1'M_2X_1)^{-1}$

$Var\left(\hat{\beta}_{1,2}\middle|X\right) \geq Var\left(\hat{\beta}_1\middle|X\right)$ （3.20）

3) $e = MY = M(X_1\beta_1 + \varepsilon) = M\varepsilon \qquad \left(M = I - X(X'X)^{-1}X',\ \ X = (X_1, X_2)\right)$

$\because MX = 0 \Rightarrow M(X_1, X_2) = 0 \Rightarrow MX_1 = 0$

$\therefore E\left(e'e\middle|X\right) = \sigma^2(n-K) \Rightarrow E\left(\frac{e'e}{n-K}\middle|X\right) = \sigma^2$ (Unbiased) （3.21）

4) $Est.Var\left(\hat{\beta}_{1,2}\middle|X\right) \geq Est.Var\left(\hat{\beta}_1\middle|X\right).$ （3.22）

5) Extensions

① $\hat{\sigma}_1^2 = e_1'e_1/(n-K_1),\ \ \hat{\sigma}_2^2 = e_2'e_2/(n-K)$

$$\frac{(n-K_1)\hat{\sigma}_1^2}{\sigma^2} = \varepsilon'M_1\varepsilon/\sigma^2 \sim \chi^2(n-K_1)$$

$$\frac{(n-K)\hat{\sigma}_2^2}{\sigma^2} = \varepsilon' M \varepsilon / \sigma^2 \sim \chi^2(n-K)$$

$$\begin{cases} Var\left(\left.\frac{(n-K_1)\hat{\sigma}_1^2}{\sigma^2}\right| X\right) = 2(n-K_1) \Rightarrow Var\left(\hat{\sigma}_1^2 \middle| X\right) = 2\sigma^4/(n-K_1) \\ Var\left(\left.\frac{(n-K)\hat{\sigma}_2^2}{\sigma^2}\right| X\right) = 2(n-K) \Rightarrow Var\left(\hat{\sigma}_2^2 \middle| X\right) = 2\sigma^4/(n-K) \end{cases}$$

$$\Rightarrow Var\left(\hat{\sigma}_1^2 \middle| X\right) < Var\left(\hat{\sigma}_2^2 \middle| X\right) \qquad (3.23)$$

In the specified model, estimated variance is more volatile.

② **t-test**

In the correct model: $$\frac{\hat{\beta}_1^{(j)}}{\sqrt{\left(Est.Var\left(\hat{\beta}_1 \middle| X\right)\right)_{jj}}} \sim t(n-K_1) \qquad (3.24)$$

In the specified model: $$\frac{\hat{\beta}_{1,2}^{(j)}}{\sqrt{\left(Est.Var\left(\hat{\beta}_{1,2} \middle| X\right)\right)_{jj}}} \sim t(n-K) \qquad (3.25)$$

Since $n-K < n-K_1$, the t distribution has a fatter tail for $\frac{\hat{\beta}_{1,2}^{(j)}}{\sqrt{\left(Est.Var\left(\hat{\beta}_{1,2} \middle| X\right)\right)_{jj}}} \sim t(n-K)$. So, if t-statistic lies in $[t_{\alpha/2}(n-K_1), t_{\alpha/2}(n-K)]$, H_0 will be accepted in the specified model, while it should be rejected in the correct model, i.e. The loss of power for t-test.

6) Conclusion:

① Exclusion of Relevant Variables

(i). biased estimator

(ii). incorrect inference procedure for t, F

② Inclusion of Irrelevant Variables

(i). loss of efficiency

The specified model loses efficiency since $E\left(\hat{\beta}_1 \middle| X\right) = \beta_1$, while

$Var\left(\hat{\beta}_{1,2}\middle|X\right)$ increase except that $X_1'X_2=0$, if $X_1'X_2=0$, then we have $X_1'X_1=X_1'M_2X_1,\ Var\left(\hat{\beta}_1\middle|X\right)=Var\left(\hat{\beta}_{1,2}\middle|X\right)$.

(ii). loss of testing power except that $X_1'X_2=0$.

It would seem that one would generally want to "overfit" the model. From a theoretical standpoint, the difficulty with this view is that the failure to use correct information is always costly. In this instance, the cost is the reduced precision of the estimates. As we have shown, the covariance matrix in the short regression (omitting X_2) is never large than the covariance matrix for the estimator obtained in the presence of superfluous variables.

3.2.3 A More General Test of Specification Error

The Ramsey **RESET test** (Regression Specification Error Test)
Ramsey has argued that various specification errors listed above (omission of relevant variables, incorrect functional form, correlation between X and ε) give rise to a nonzero ε vector.

H_0 :: no specification error

Augmented Regression: $Y=X\beta+Z\alpha+\varepsilon$. （3.26）

Where $Z=\left(\hat{Y}^2,\hat{Y}^3,\cdots\right)$

Under the null of no specification error, the coefficients of Z are zero. RESET is simple a Wald test of $\alpha=0$, but its power is limited**.**

3.3 Nonlinearity

3.3.1 Intrinsic Linearity (P_{122},$P_{128\text{-}129}$,Greene^{5th})

Let $z = z_1, z_2, \cdots, z_L$ be a set of L independent variables ; Let $f_1, f_2, \cdots, f_K$ be K linearly independent function of z ; Let $g(y)$ be an observable function of y ; and retain the usual assumptions about the disturbance. The linear regression model is

$$\begin{aligned} g(y) &= \beta_1 f_1(z) + \beta_2 f_2(z) + \cdots + \beta_K f_K(z) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \\ &= x'\beta + \varepsilon \end{aligned} \tag{3.27}$$

In the classical linear regression model, if the K parameters $\beta_1, \beta_2, \cdots, \beta_K$ can be written as K one-to-one, possibly nonlinearity functions of a set of K underlying parameters $\theta_1, \theta_2, \cdots, \theta_K$, then the model is intrinsically linear in θ.

The one-to-one correspondence is an identification condition. If the condition is met, then the underlying parameters of the regression (θ) are said to be exactly identified in terms of the parameters of the linear model β.

1) Example 7.5 CES Production Function(5th P129)

The constant elasticity of substitution production function may be written

$$Y = \gamma\left[\delta K^{-\rho} + (1-\delta) L^{-\rho}\right]^{-\frac{\nu}{\rho}} e^{\varepsilon} \tag{3.28}$$

$$\ln Y = \ln\gamma - \frac{\nu}{\rho}\ln\left[\delta K^{-\rho} + (1-\delta) L^{-\rho}\right] + \varepsilon$$

Denotes $f(\rho) = \frac{1}{\rho}\ln\left[\delta K^{-\rho} + (1-\delta) L^{-\rho}\right] = \frac{\ln\left[g(\rho)\right]}{\rho}$, use Taylor series approximation to $f(\rho)$ around the point $\rho = 0$.

$$f(\rho) \approx f(\rho)\big|_{\rho=0} + (\rho - 0) f'(\rho)\big|_{\rho=0}$$

$$f(\rho)\big|_{\rho=0} = \left.\frac{\ln\left[g(\rho)\right]}{\rho}\right|_{\rho=0} = \left.\frac{g'(\rho)}{g(\rho)}\right|_{\rho=0}$$

$$= \left(-\delta K^{-\rho}\ln K - (1-\delta) L^{-\rho}\ln L\right)\Big|_{\rho=0} = -\delta\ln K - (1-\delta)\ln L$$

$$f'(\rho)\Big|_{\rho=0} = \left.\frac{\frac{g'(\rho)}{g(\rho)}\rho - \ln\left[g(\rho)\right]}{\rho^2}\right|_{\rho=0} = \left.\frac{\frac{g''(\rho)g(\rho)-\left[g'(\rho)\right]^2}{\left[g(\rho)\right]^2}\rho}{2\rho}\right|_{\rho=0}$$

$$= \left.\frac{g''(\rho)g(\rho)-\left[g'(\rho)\right]^2}{2\left[g(\rho)\right]^2}\right|_{\rho=0}$$

$$= \frac{1}{2}\left.\left\{\left[\delta K^{-\rho}(\ln K)^2 + (1-\delta)L^{-\rho}(\ln L)^2\right] - \left[-\delta K^{-\rho}\ln K - (1-\delta)L^{-\rho}\ln L\right]^2\right\}\right|_{\rho=0}$$

$$= \frac{1}{2}\left\{\left[\delta(\ln K)^2 + (1-\delta)(\ln L)^2\right] - \left[-\delta\ln K - (1-\delta)\ln L\right]^2\right\}$$

$$= \frac{1}{2}\delta(1-\delta)(\ln K - \ln L)^2$$

$$\therefore f(\rho) \approx f(\rho)\Big|_{\rho=0} + (\rho-0)f'(\rho)\Big|_{\rho=0}$$

$$= -\delta\ln K - (1-\delta)\ln L + \frac{1}{2}\rho\delta(1-\delta)(\ln K - \ln L)^2$$

$$\therefore \ln Y = \ln\gamma - \frac{\nu}{\rho}\ln\left[\delta K^{-\rho} + (1-\delta)L^{-\rho}\right] + \varepsilon$$

$$= \ln\gamma + \nu\delta\ln K + \nu(1-\delta)\ln L - \nu\rho\delta(1-\delta)\left[\frac{1}{2}(\ln K - \ln L)^2\right] + \varepsilon \qquad (3.29)$$

$$= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Where $x_1 = 1$, $x_2 = \ln K$, $x_3 = \ln L$, $x_4 = -\frac{1}{2}(\ln K - \ln L)^2$, and the transformations are

$\beta_1 = \ln\gamma$, $\beta_2 = \nu\delta$, $\beta_3 = \nu(1-\delta)$, $\beta_4 = \nu\rho\delta(1-\delta)$

So $\gamma = e^{\beta_1}$, $\delta = \beta_2/(\beta_2+\beta_3)$, $\nu = \beta_2+\beta_3$, $\rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3)$ （3.30）

We would use the delta method to construct the estimated asymptotic covariance matrix for the estimates of $\theta' = [\gamma, \delta, \nu, \rho]$. The derivatives matrix is

$$C = \frac{\partial\theta}{\partial\beta'} = \begin{bmatrix} e^{\beta_1} & 0 & 0 & 0 \\ 0 & \beta_3/(\beta_2+\beta_3)^2 & -\beta_2/(\beta_2+\beta_3)^2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -\beta_3\beta_4/(\beta_2^2\beta_3) & -\beta_2\beta_4/(\beta_2\beta_3^2) & (\beta_2+\beta_3)/(\beta_2\beta_3) \end{bmatrix} \qquad (3.31)$$

The estimated covariance matrix for $\hat{\theta}$ is $\hat{C}\left[s^2(X'X)^{-1}\right]\hat{C}'$.

2) Other examples of intrinsic linear models

① **Log_linear**

$$\begin{cases} \ln Y = \beta_1 + \sum \beta_k \ln X_k + \varepsilon \quad where\ \beta_k = \dfrac{\partial \ln Y}{\partial \ln X_k} \to elasticity \to percentage\ change \\ Y = \beta_1 + \sum \beta_k X_k + \varepsilon \to \beta_k = \dfrac{dY}{dX_k} \to absolute\ change \end{cases}$$

② **Semilog_growth**

$$\ln Y = \beta_1 + \beta_2 t + \varepsilon \Rightarrow \beta_2 = \frac{d \ln Y}{dt} \to growth\ rate$$

$\ln \dfrac{Q}{L} = \beta_1 + \beta_2 \ln \dfrac{K}{L} + \delta t + \varepsilon_t \to \delta$: average growth rate of $\dfrac{Q}{L}$ while keep the use of $\dfrac{K}{L}$ constant.

$$Y = \beta_1 + \beta_2 \ln X + \varepsilon \Rightarrow \beta_2 = \frac{dY}{d \ln X}$$

Sometimes, the effect of the variable is measured by percentage change. For example, it can reflect the effect on output level Y if money supply increases by 1%.

③ **Interaction terms**

$$Y = \beta_1 + \beta_2 HR + \beta_3 FDI + \beta_4 FDI \cdot HR + \varepsilon$$

$$\frac{\partial Y}{\partial FDI} = \beta_3 + \beta_4 HR$$

HR denotes Human Resources.

④**Square terms**

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_4 X_2^3 + \varepsilon \Rightarrow \frac{\partial Y}{\partial X_2} = \beta_2 + 2\beta_3 X_2 + 3\beta_4 X_2^2$$

Linear function of X_2 is not true especially with nonconstant decreasing or increasing rate. e.g. Taxation revenue is a quadratic function of taxation ratio.

$$Wage = \beta_1 + \beta_2 age + \beta_3 age^2 + \varepsilon \qquad (\beta_3 < 0)$$

for young age 30-40: $\dfrac{\partial Wage}{\partial age} = \beta_2 + 2\beta_3 age$

for old age 50-60: $\dfrac{\partial Wage}{\partial age} = \beta_2 + 2\beta_3 age$

The latter $\dfrac{\partial Wage}{\partial age}$ is less than the former one.

3.3.2 Nonlinear Regression Model

$y_i = h\left(x_i, \beta^0\right) + \varepsilon_i$

e.g. $y_i = \beta_1 + \beta_2 e^{\beta_3 x_i} + \varepsilon_i$

$S_n(\beta) = \sum_{i=1}^{n}\left[y_i - h(x_i, \beta)\right]^2$

FOC:

$$\begin{aligned}
\frac{\partial S_n(\beta)}{\partial \beta_1} &= -\sum_{i=1}^{n}\left(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i}\right) = 0 \\
\frac{\partial S_n(\beta)}{\partial \beta_2} &= -\sum_{i=1}^{n}\left(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i}\right) e^{\beta_3 x_i} = 0 \\
\frac{\partial S_n(\beta)}{\partial \beta_3} &= -\sum_{i=1}^{n}\left(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i}\right) \beta_2 e^{\beta_3 x_i} = 0
\end{aligned} \tag{3.32}$$

The equations (3.32) have no closed-form solution.
A nonlinear regression model is the one for which the first-order condition for least squares estimation of the parameters are nonlinear functions of the parameters.

3.4 Choosing between Nonnested Models

The classical testing procedures that we have been using have been shown to be most powerful for the types of hypotheses we have considered. Although use of these procedures is clearly desirable, the requirement that we express the hypotheses in the form of the restrictions on the model $Y = X\beta + \varepsilon$,

$H_0: R\beta = q$ versus $H_1: R\beta \neq q$, can be limiting.

There is one common exception: which of two possible sets of regressors is more appropriate?

3.4.1 An Encompassing Model (P_{154}, Greene^{5th})

$$H_0: Y = X\beta + \varepsilon \quad H_1: Y = Z\gamma + \varepsilon \tag{3.33}$$

An artificial nesting of the two models, i.e. "Supermodel":

$$Y = \overline{X}\,\overline{\beta} + \overline{Z}\,\overline{\gamma} + W\delta + \varepsilon \tag{3.34}$$

$\overline{X}$: the set of variables in X that are not in Z,

$\bar{Z}$: the set of variables in Z that are not in X ,

W : the variables that the two models have in common.

F test: $H_0 : \bar{\gamma} = 0$

Accept $H_0 \Rightarrow Model\ Y = X\beta + \varepsilon$

Reject $H_0 \Rightarrow Model\ Y = \bar{X}\bar{\beta} + \bar{Z}\bar{\gamma} + W\delta + \varepsilon$

Problems:

(i). δ remains a mixture of parts of β and γ , this test does not really distinguish between H_0 and H_1;

(ii). In a time-series setting, the problem of collinearity may be severe.

3.4.2 The J Test (Davidson & MacKinnon, 1981)

Model: $Y = (1-\alpha)X\beta + \alpha Z\gamma + \varepsilon$ (3.35)

$H_0 : \alpha = 0 \ \ (against\ H_1 : Y = Z\gamma + \varepsilon_1)$

$H_1 : \alpha = 1 \ \ (against\ H_0 : Y = X\beta + \varepsilon)$

Problems

(i). nonlinearity

(ii). α can not be separately estimated, we can only get $(1-\alpha)\beta$ & $\alpha\gamma$.

J Test: $t = \hat{\alpha}/se(\hat{\alpha}) \overset{d}{\sim} N(0,1)$ (3.36)

① X is the appropriate set of variables to explain Y

a. $Y = Z\gamma + \varepsilon_1 \Rightarrow \hat{Y}^{<1>} = Z\hat{\gamma}$ (3.37)

b. $Y = X\beta + \lambda^{<1>}\hat{Y}^{<1>} + \varepsilon \Rightarrow \hat{\lambda}^{<1>}$ (3.38)

c. Testing the significance of $\lambda^{<1>}$.

Accept $\lambda^{<1>} = 0$, $\Rightarrow$ accept $H_0 : Y = X\beta + \varepsilon$, Z can not help to explain Y .

Reject $\lambda^{<1>} = 0$, $\Rightarrow$ accept $H_1 : Y = Z\gamma + \varepsilon$, Z can significantly help to explain Y .

② Z is the appropriate set of variables to explain Y

a. $H_0: Y = X\beta + \varepsilon_0 \Rightarrow \hat{Y}^{<0>} = X\hat{\beta}$ (3.39)

b. $Y = Z\gamma + \hat{Y}^{<0>}\lambda^{<0>} + \varepsilon_1 \Rightarrow \hat{\lambda}^{<0>}$ (3.40)

c. Testing the significance of $\lambda^{<0>}$

Accept $\lambda^{<0>} = 0$, $\Rightarrow$ Accept H_1: X can not help to explain Y.

Reject $\lambda^{<0>} = 0$, $\Rightarrow$ Accept H_0: X can significantly help to explain Y.

③ Accept both of H_0 and H_1

The data is not sufficient to distinguish between H_0 and H_1.

④ Reject both: Two sets of variables can not help to explain Y.

Example:

$H_0: C_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_{t-1} + \varepsilon_{0t}$

$H_1: C_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 C_{t-1} + \varepsilon_{1t}$

H_0: consumption responds to changes in income over two periods

H_1: consumption responds to changes in income over many periods

3.4.3 Cox Test (1961,1962)

1) H_0: X is the appropriate set of variables

$$Q = \frac{c_{01}}{\sqrt{v_{01}}} \overset{d}{\sim} N(0,1), \tag{3.41}$$

$$v_{01} = \frac{s_X^2 \cdot b'X'M_Z M_X M_Z Xb}{s_{ZX}^4} \tag{3.42}$$

$$c_{01} = \frac{n}{2}\ln\left[\frac{s_Z^2}{s_{ZX}^2}\right] = \frac{n}{2}\ln\left[\frac{s_Z^2}{s_X^2 + (1/n)b'X'M_Z Xb}\right] \tag{3.43}$$

$$M_Z = I - Z(Z'Z)^{-1}Z'$$

$$M_X = I - X(X'X)^{-1}Z'$$

$$b = (X'X)^{-1}X'Y$$

$s_Z^2 = \frac{1}{n} e_Z' e_Z$: mean-squared residual in the regression of Y on Z

$s_X^2 = \frac{1}{n} e_X' e_X$: mean-squared residual in the regression of Y on X

$s_{ZX}^2 = s_X^2 + \frac{1}{n} b' X' M_Z X b$

2) H_0: Z is the appropriate set of variables

$$q = \frac{c_{10}}{\sqrt{v_{10}}} \overset{d}{\sim} N(0,1) \quad (3.43)$$

$$v_{10} = s_Z^2 \cdot \frac{d' Z' M_X M_Z M_X Z d}{s_{XZ}^4} \quad (3.44)$$

$$c_{10} = \frac{n}{2} \ln [\frac{s_X^2}{s_{XZ}^2}] \quad (3.45)$$

$s_{XZ}^2 = s_Z^2 + \frac{1}{n} d' Z' M_X Z d$