# 5. NONLINEAR MODELS

## [1] Nonlinear (NL) Regression Models

- General form of nonlinear or linear regression models:

$$y_t = h(x_{t\bullet},\beta) + \varepsilon_t, \quad \varepsilon_t \text{ iid } N(0,\sigma^2).$$

- Assume that the $x_{t\bullet}$ and $\varepsilon_t$ stochastically independent.

  → This assumption implies that:

$$E\left(\frac{\partial h(x_{t\bullet},\beta)}{\partial\beta}\varepsilon_t\right) = 0_{k\times 1};$$

$$E\left(\frac{\partial^2 h(x_{t\bullet},\beta)}{\partial\beta\partial\beta'}\varepsilon_t\right) = 0_{k\times k}.$$

Example 1:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t.$$

Example 2:

$y_t = \beta_1 + \beta_2 e^{\beta_3 x_t} + \varepsilon_t$, where $x_t$ is a scalar.

Example 3:

$$y_t = Ax_{t2}^{\beta_2}x_{t3}^{\beta_3} + \varepsilon_t. \qquad (*)$$

## [2] Estimation of NL models

**Definition:**

Let $S(\beta) = \Sigma_t \left( y_t - h(x_{t\bullet}, \beta) \right)^2$. The NLLS (Nonlinear least squares) estimator, $\hat{\beta}_{NL}$, minimizes S(β).

**Assumptions:** See Greene, Chapter 9.

**Facts:**

1) $p\lim \hat{\beta}_{NL} = \beta_o$ (consistent). [No guarantee that $E\left(\hat{\beta}_{NL}\right) = \beta_o$.]

2) $\sqrt{T}(\hat{\beta}_{NL} - \beta_o) \rightarrow_d N\left( 0_{k\times 1}, p\lim \sigma^2 \left[ \frac{1}{T} \Sigma_t H_t(\beta_o) H_t(\beta_o)' \right]^{-1} \right)$,

where $H_t(\beta) = \partial h(x_{t\bullet}, \beta) / \partial \beta$.

If we define $H(\beta) = \begin{pmatrix} H_1(\beta)' \\ H_2(\beta)' \\ \vdots \\ H_T(\beta)' \end{pmatrix}$, $H(\beta)'H(\beta) = \Sigma_t H_t(\beta) H_t(\beta)'$.

3) Let $s_{NL}^2 = S(\hat{\beta}_{NL})/T$, Then, $p\lim_{T\rightarrow\infty} s_{NL}^2 = \sigma^2$.

4) $\hat{\beta}_{NL} \approx N\left( \beta, s_{NL}^2 \left[ H(\hat{\beta}_{NL})' H(\hat{\beta}_{NL}) \right]^{-1} \right)$.

5) $R^2$ = 1 - SSE/SST, where $SSE = S(\hat{\beta}_{NL})$.

→ There is no guarantee that $0 \le R^2 \le 1$.

Example:

$$y_t = \beta_1 + \beta_2 e^{\beta_3 x_t} + \varepsilon_t .$$

Then,

$$\frac{\partial h(x_{t\bullet}, \beta)}{\partial \beta_1} = 1; \frac{\partial h(x_{t\bullet}, \beta)}{\partial \beta_2} = e^{\beta_3 x_t}; \frac{\partial h(x_{t\bullet}, \beta)}{\partial \beta_3} = \beta_2 x_t e^{\beta_3 x_t} .$$

Thus, $H_t(\beta) = \begin{pmatrix} 1 \\ e^{\beta_3 x_t} \\ \beta_2 x_t e^{\beta_3 x_t} \end{pmatrix}$.

<Sketchy Proof of Consistency>

- From $\partial S(\hat{\beta}_{NL}) / \partial \beta = 0_{k\times 1}$,

$$\Sigma_{t=1}^{T} H_t(\hat{\beta}_{NL})(y_t - h_t(\hat{\beta}_{NL})) = 0_{k\times 1} .$$

- Taylor expansion around β = β$_o$:

$$0_{k\times 1} \approx \Sigma_t H_t(\beta_o)(y_t - h_t(\beta_o))$$
$$+\Sigma_t \left( \frac{\partial^2 h_t(\beta_o)}{\partial \beta \partial \beta'} \left( y_t - h_t(\beta_o) \right) - H_t(\beta_o) H_t(\beta_o)' \right) \left( \hat{\beta}_{NL} - \beta_o \right).$$

→

$$\left( \hat{\beta}_{NL} - \beta_o \right) \approx - \left( \Sigma_t \left( \frac{\partial^2 h_t(\beta_o)}{\partial \beta \partial \beta'} \left( y_t - h_t(\beta_o) \right) - H_t(\beta_o) H_t(\beta_o)' \right) \right)^{-1}$$
$$\times \Sigma_t H_t(\beta_o)(y_t - h_t(\beta_o)).$$

→

$$\left(\hat{\beta}_{NL}-\beta_o\right)\approx-\left(\frac{1}{T}\Sigma_t\left(\frac{\partial^2 h_t(\beta_o)}{\partial\beta\partial\beta'}\left(y_t-h_t(\beta_o)\right)-H_t(\beta_o)H_t(\beta_o)'\right)\right)^{-1}$$

$$\times\frac{1}{T}\Sigma_t H_t(\beta_o)(y_t-h_t(\beta_o))$$

$$\approx-\left(\frac{1}{T}\Sigma_t\left(\frac{\partial^2 h_t(\beta_o)}{\partial\beta\partial\beta'}\varepsilon_t-H_t(\beta_o)H_t(\beta_o)'\right)\right)^{-1}\frac{1}{T}\Sigma_t H_t(\beta_o)\varepsilon_t$$

$$\approx\left(\frac{1}{T}\Sigma_t H_t(\beta_o)H_t(\beta_o)'\right)^{-1}\times 0_{k\times 1}=0_{k\times 1}.$$

<Sketchy Proof of Asymptotic Normality>

- Note that

$$\sqrt{T}\left(\hat{\beta}_{NL}-\beta_o\right)\approx-\left(\frac{1}{T}\Sigma_t\left(\frac{\partial^2 h_t(\beta_o)}{\partial\beta\partial\beta'}\left(y_t-h_t(\beta_o)\right)-H_t(\beta_o)H_t(\beta_o)\right)\right)^{-1}$$

$$\times\frac{1}{\sqrt{T}}\Sigma_t H_t(\beta_o)(y_t-h_t(\beta_o))$$

$$\approx\left(\frac{1}{T}\Sigma_t H_t(\beta_o)H_t(\beta_o)'\right)^{-1}\frac{1}{\sqrt{T}}\Sigma_t H_t(\beta_o)(y_t-h_t(\beta_o)).$$

- But

$$\frac{1}{\sqrt{T}}\Sigma_t H_t(\beta_o)(y_t-h_t(\beta_o))=\frac{1}{\sqrt{T}}\Sigma_t H_t(\beta_o)\varepsilon_t$$

$$\to_d N\left(0_{k\times 1}, p\lim\sigma^2\frac{1}{T}\Sigma_t H_t(\beta_o)H_t(\beta_o)'\right).$$

## [3] SPECIFICATION TESTS

### CASE A:

- Same dependent variables under both the null and the alternatives

$$H_o: y_t = h^o(x_{t\bullet},\beta) + \varepsilon_t. \qquad (A)$$

$$H_a: y_t = h^a(w_{t\bullet},\gamma) + \varepsilon_t. \qquad (B)$$

Example:

$h^o(x_{t\bullet},\beta) = \beta_1 + \beta_2 x_t$ ; $h^a(w_{t\bullet},\gamma) = \gamma_1 + \gamma_2 \ln(x_t)$.

[Even if the models are linear, we still can perform J or P tests.]

(1) J Test: Davidson and Mackinnon (1981, Econometrica)

- Construct the following auxiliary model:

$$y_t = (1-\alpha)h^o(x_{t\bullet},\beta) + \alpha h^a(w_{t\bullet},\gamma) + \varepsilon_t. \qquad (C)$$

- If $H_o$ is correct, $\alpha = 0$.
- Let $\hat{\gamma}_{NL}$ be the NLLS estimator of $\gamma$ from (B). Replace $h^a(w_{t\bullet},\gamma)$ by $\hat{h}_t^a = h^a(w_{t\bullet},\hat{\gamma}_{NL})$ (fitted value of $y_t$ from (B)):

$$y_t = (1-\alpha)h^o(x_{t\bullet},\beta) + \alpha \hat{h}_t^a + error. \qquad (D)$$

- Do NLLS on (D), and estimate $\beta$ and $\alpha$ jointly. Using the estimates, we can perform a t-test for $H_o'$: $\alpha = 0$. [In the sense that we estimate $\beta$ and $\alpha$ jointly, we call the test J-test.]

(2) P-Test: An alternative to J-test.

- Get $\hat{\beta}_{NL}$ and $\hat{\gamma}_{NL}$ by NLLS on both (A) and (B).
- Consider the following auxiliary regression:

$$y_t - \hat{h}_t^o = \hat{H}_t^{o\prime} b + (\hat{h}_t^a - \hat{h}_t^o)\alpha + error, \quad \text{(E)}$$

where $\hat{h}_t^o = h^o(x_{t\bullet}, \hat{\beta}_{NL})$, $\hat{h}_t^a = h^a(w_{t\bullet}, \hat{\gamma}_{NL})$ and

$\hat{H}_t^o = H_t^o(x_{t\bullet}, \hat{\beta}_{NL})$.

- Do OLS on (E) and estimate b and $\alpha$. Then, perform t-test for $H_o'$: $\alpha = 0$.

**CASE B:**

- Different dependent variables

$H_o$: $y_t = h^o(x_{t\bullet},\beta) + \varepsilon_t$.

$H_a$: $g(y_t) = h^a(w_{t\bullet},\gamma) + \varepsilon_t$,

where $g(y_t)$ is a function of $y_t$ (e.g., $g(y_t) = \ln(y_t)$).

Example:

$H_o$: $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$

$H_a$: $\ln(y_t) = \gamma_1 + \gamma_2 \ln(x_t) + \varepsilon_t$

- P test:

  Estimate both β and γ by NLLS (or OLS).

  Construct the following auxiliary model:

$$y_t - \hat{h}_t^o = \hat{H}_t^o b + (\hat{h}_t^a - g(\hat{h}_t^o))\alpha + error . \qquad \text{(F)}$$

  Do OLS on (F), and test $H_o'$: $\alpha = 0$.

Example:

$H_o$: $y_t = x_{t\bullet}'\beta + \varepsilon_t$

$H_a$: $\ln(y_t) = w_{t\bullet}'\gamma + \varepsilon_t$ ,

where $x_{t\bullet} = [1,x_{t2},...,x_{tk}]'$ and $w_{t\bullet} = [1,\ln(x_{t2}),...,\ln(x_{tk})]$.

Explain in detail how you would test $H_o$.

[EXAMPLE]

- Data: (WAGE2.WF1 or WAGE2.TXT – from Wooldridge's website)

  # of observations (T): 935

| | |
|---|---|
| 1. wage | monthly earnings |
| 2. hours | average weekly hours |
| 3. IQ | IQ score |
| 4. KWW | knowledge of world work score |
| 5. educ | years of education |
| 6. exper | years of work experience |
| 7. tenure | years with current employer |
| 8. age | age in years |
| 9. married | =1 if married |
| 10. black | =1 if black |
| 11. south | =1 if live in south |
| 12. urban | =1 if live in SMSA |
| 13. sibs | number of siblings |
| 14. brthord | birth order |
| 15. meduc | mother's education |
| 16. feduc | father's education |
| 17. lwage | natural log of wage |

$H_o$: $lwage_t = \beta_1 + \beta_2 educ_t + \beta_3 exper_t + \varepsilon_t$.

$H_a$: $lwage_t = \gamma_1 + \gamma_2 \left( educ_t^{\gamma_3} + exper_t^{\gamma_4} \right) + \varepsilon_t$.

<J-Test>

= STEP 1 =

Estimate the model under $H_a$:

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935
Estimation settings: tol= 0.00010, derivs=analytic
Initial Values: C(1)=1.00000, C(2)=1.00000, C(3)=1.00000, C(4)=1.00000
Convergence achieved after 80 iterations
LWAGE=C(1)+C(2)*(EDUC^C(3)+EXPER^C(4))

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 5.667911 | 0.444198 | 12.75986 | 0.0000 |
| C(2) | 0.035137 | 0.054108 | 0.649384 | 0.5163 |
| C(3) | 1.221412 | 0.446902 | 2.733068 | 0.0064 |
| C(4) | 0.832149 | 0.431536 | 1.928344 | 0.0541 |

| | | | |
|---|---|---|---|
| R-squared | 0.129527 | Mean dependent var | 6.779004 |
| Adjusted R-squared | 0.126723 | S.D. dependent var | 0.421144 |
| S.E. of regression | 0.393556 | Akaike info criterion | 0.977083 |
| Sum squared resid | 144.1993 | Schwarz criterion | 0.997791 |
| Log likelihood | -452.7862 | Durbin-Watson stat | 1.786041 |

Get fitya = lwage - resid

= STEP 2 =

Estimate $lwage = (1-\alpha)(\beta_1 + \beta_2 educ + \beta_3 exer) + \alpha\, fitya + error$.

Dependent Variable: LWAGE
Method: Least Squares
Date: 04/09/02 Time: 12:26
Sample: 1 935
Included observations: 935
Estimation settings: tol= 0.00010, derivs=analytic
Initial Values: C(1)=5.66791, C(2)=0.03514, C(3)=1.22141,
C(4)=0.83215
Convergence achieved after 22 iterations
LWAGE=(1-C(1))*(C(2)+C(3)*EDUC+C(4)*EXPER)+C(1)*FITYA

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| **C(1)** | **-4.692962** | **2.699983** | **-1.738145** | **0.0825** |
| C(2) | 5.507546 | 0.019665 | 280.0728 | 0.0000 |
| C(3) | 0.077409 | 0.001155 | 67.04214 | 0.0000 |
| C(4) | 0.019793 | 0.000579 | 34.15696 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.133671 | Mean dependent var | 6.779004 |
| Adjusted R-squared | 0.130879 | S.D. dependent var | 0.421144 |
| S.E. of regression | 0.392618 | Akaike info criterion | 0.972312 |
| Sum squared resid | 143.5129 | Schwarz criterion | 0.993020 |
| Log likelihood | -450.5558 | Durbin-Watson stat | 1.793672 |

Do not reject $H_o$: C(1) = 0 ($\alpha = 0$).

<P-Test>

= STEP 1 =

Estimate the model under $H_o$:

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935
LWAGE=C(1)+C(2)*EDUC+C(3)*EXPER

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 5.502710 | 0.112037 | 49.11510 | 0.0000 |
| C(2) | 0.077782 | 0.006577 | 11.82660 | 0.0000 |
| C(3) | 0.019777 | 0.003303 | 5.988419 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.130859 | Mean dependent var | 6.779004 |
| Adjusted R-squared | 0.128994 | S.D. dependent var | 0.421144 |
| S.E. of regression | 0.393044 | Akaike info criterion | 0.973413 |
| Sum squared resid | 143.9786 | Schwarz criterion | 0.988944 |
| Log likelihood | -452.0704 | Durbin-Watson stat | 1.787824 |

Get fity0 = lwage – resid and res0 = resid.

= STEP 2 =

Estimate the model under $H_a$:

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935
Estimation settings: tol= 0.00010, derivs=analytic
Initial Values: C(1)=1.00000, C(2)=1.00000, C(3)=1.00000,
C(4)=1.00000
Convergence achieved after 80 iterations
LWAGE=C(1)+C(2)*(EDUC^C(3)+EXPER^C(4))

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 5.667911 | 0.444198 | 12.75986 | 0.0000 |
| C(2) | 0.035137 | 0.054108 | 0.649384 | 0.5163 |
| C(3) | 1.221412 | 0.446902 | 2.733068 | 0.0064 |
| C(4) | 0.832149 | 0.431536 | 1.928344 | 0.0541 |

| | | | |
|---|---|---|---|
| R-squared | 0.129527 | Mean dependent var | 6.779004 |
| Adjusted R-squared | 0.126723 | S.D. dependent var | 0.421144 |
| S.E. of regression | 0.393556 | Akaike info criterion | 0.977083 |
| Sum squared resid | 144.1993 | Schwarz criterion | 0.997791 |
| Log likelihood | -452.7862 | Durbin-Watson stat | 1.786041 |

Get fitya = lwage – resid.

= STEP 3 =

Estimate $res0 = \beta_1 + \beta_2 educ + \beta_3 \exp er + \alpha(fitya - fity0)$.

Dependent Variable: RES0
Method: Least Squares
Sample: 1 935
Included observations: 935
RES0=C(1)+C(2)*EDUC+C(3)*EXPER+C(4)*(FITYA-FITY0)

| | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 0.027535 | 0.113031 | 0.243602 | 0.8076 |
| C(2) | -0.002125 | 0.006682 | -0.317934 | 0.7506 |
| C(3) | 9.34E-05 | 0.003299 | 0.028313 | 0.9774 |
| **C(4)** | **-4.692962** | **2.699983** | **-1.738145** | **0.0825** |

| | | | |
|---|---|---|---|
| R-squared | 0.003235 | Mean dependent var | 3.38E-15 |
| Adjusted R-squared | 0.000023 | S.D. dependent var | 0.392623 |
| S.E. of regression | 0.392618 | Akaike info criterion | 0.972312 |
| Sum squared resid | 143.5129 | Schwarz criterion | 0.993020 |
| Log likelihood | -450.5558 | Durbin-Watson stat | 1.793672 |

Do not reject $H_o$ at 5% of significance level.

## [4] NL models with unknown parameters in LHS

- General form: $g(y_t,\delta) = h(x_{t\bullet},\beta) + \varepsilon_t$, $\varepsilon_t$ iid $N(0,\sigma^2)$.

Example: Generalized Cobb-Douglas Function

$$\ln(y_t) + \delta y_t = \beta_1 + \beta_2(1-\beta_3)\ln(K_t) + \beta_2\beta_3\ln(L_t) + \varepsilon_t.$$

If $\delta = 0$, the function becomes Cobb-Douglas.

Estimation:

1) NLLS could be inconsistent. Even if it is consistent, computation of the covariance matrix of the NLLS estimator could be complicated.
2) Do MLE. (See Greene.)

Example: Box-Cox transformation

- Define: $y_t(\delta) = [y_t^{\delta} - 1]/\delta$ ; $x_{tj}(\lambda) = [x_{tj}^{\lambda} - 1]/\lambda$, $j = 2, \ldots , k$.
- Assume that $\lambda$ is all the same for j, we can allow $\lambda$ to vary over different j.
- If $\lambda = 1$, $x_{tj}(1) = x_{tj} - 1$ (linear). If $\lambda \to 0$, $x_{tj}(\lambda) \to \ln(x_{tj})$ (log).

- Box-Cox Model:
  - $y_t(\delta) = \beta_1 + \Sigma_{j=2}^{k} x_{tj}(\lambda)\beta_j + \varepsilon_t$.
  - $y_t(\delta) = x_{t\bullet}(\lambda)'\beta + \varepsilon_t$,

    where $x_{t\bullet}(\lambda) = [1, x_{t2}(\lambda), ..., x_{tk}(\lambda)]'$ and $\beta = [\beta_1, ..., \beta_k]'$.
  - Use MLE to estimate $\delta$, $\lambda$ and $\beta$.