# 非線形関数の回帰 (1)

別所俊一郎

**2006 年 5 月 31 日**

## Today's attraction

- 重回帰分析の非線形関数への応用の概論
- 一般的な手順
- 独立変数が１つのときの非線形関数の代表例
- 多項式関数と対数関数の比較

# ここで扱う非線形関数

非線形関数では，説明変数 $X_1$ の変化がもたらす被説明変数 $Y$ の変化が一定ではないが，ここではとりわけ以下の 2 つのケースを扱う

- 説明変数 $X_1$ の変化がもたらす被説明変数 $Y$ の変化が $X_1$ の大きさに依存：傾きが変化
- 説明変数 $X_1$ の変化がもたらす被説明変数 $Y$ の変化が他の説明変数 $X_2$ の大きさに依存：回帰線の「位置」が $X_2$ によって変化

$E[Y_i|X_1, X_2, ..., X_k]$ が $\{X_j\}$ の非線形関数

- 未知の係数の線形関数になっているケースのみ
- 重回帰分析の応用となる
- 非線形関数の係数（推定値）それ自体の解釈は難しい
- 図を描いたり、変分を計算したりして解釈するのが望ましい。

# 非線形関数の回帰の一般的戦略（例）

テストの成績と学校区の一人当たり所得

- 強い正の相関：標本相関係数は 0.71（Fig. 6.2）
- 回帰直線を引いてみると，直線的な関係にはないようにみえる
  - 所得が高いか低いところでは，観測値は回帰直線の下方に分布
  - 所得が中程度のところでは，観測値は回帰直線の上方に分布
- 2 つの変数は直線的な関係にはない?
  - 1 つの候補は 2 次関数（quadratic function）

$$\text{点数}_i = \beta_0 + \beta_1 \times \text{所得}_i + \beta_2 \times \text{所得}_i^2 + \text{誤差項}$$

  - 係数 $\beta_0, \beta_1, \beta_2$ の真の値は未知：統計的推測
  - 非線形関数の統計的推測?

# 非線形関数の回帰の一般的戦略（例）

テストの成績と学校区の一人当たり所得

$$点数_i = \beta_0 + \beta_1 \times 所得_i + \beta_2 \times 所得_i^2 + 誤差項$$

- 2 変数の重回帰と同じ形：OLS 推定が可能
- 所得と，所得の 2 乗項を 2 つの説明変数とし，OLS 推定を行う
  - 誤差項の条件付き期待値はゼロ
  - 無作為抽出標本
  - 適切なモーメントの条件
  - 完全な多重共線性の不在
- 非線形性の検定は $H_0 : \beta_2 = 0$ を検定すればよい：棄却できない（t 検定）

# 非線形関数の回帰の一般的戦略：概論

- 被説明変数が説明変数の非線形関数であるときには，$X_1$ の変分 $\Delta X_1$ に対する $Y$ の変分 $\Delta Y$ は $\Delta Y = \beta_1 \Delta X_1$ とは限らない

- ここで扱う非線形関数は，係数と誤差項については線形で

$$Y_i = f(X_{1i}, X_{2i}, ..., X_{ki}) + u_i$$

被説明変数 $Y_i$ について条件付き期待値をとると

$$E[Y_i | X_{1i}, X_{2i}, ..., X_{ki}] = f(X_{1i}, X_{2i}, ..., X_{ki})$$

であって，$f(\cdot)$ は $(X_{1i}, X_{2i}, ..., X_{ki})$ について非線形

- $X_1$ の変分 $\Delta X_1$ に対する $Y$ の変分 $\Delta Y$ は

$$\Delta Y = f(X_1 + \Delta X_1, X_2, ..., X_k) - f(X_1, X_2, ..., X_k)$$

と定義される．実際に求められるのは，推定された係数を用いた

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, ..., X_k) - \hat{f}(X_1, X_2, ..., X_k)$$

# 非線形関数の回帰の一般的戦略（例）

テストの成績と学校区の一人当たり所得が 2 次関数の関係にあるとすると

- 所得の変化に対する成績の変化は所得水準によって異なる

$$\begin{aligned}\Delta\hat{Y}|_{X_1=10} &= (\hat{\beta}_0+\hat{\beta}_1\times 11+\hat{\beta}_2\times 11^2)-(\hat{\beta}_0+\hat{\beta}_1\times 10+\hat{\beta}_2\times 10^2)\\ &= 2.96\\ \Delta\hat{Y}|_{X_1=40} &= (\hat{\beta}_0+\hat{\beta}_1\times 41+\hat{\beta}_2\times 41^2)-(\hat{\beta}_0+\hat{\beta}_1\times 40+\hat{\beta}_2\times 40^2)\\ &= 0.42\end{aligned}$$

- Sampling error があるので、$\Delta\hat{Y}$ の標準偏差を求めれば信頼区間を形成できる
  - $X_1$ が $10\to 11$ のとき、$\Delta\hat{Y}=\hat{\beta}_1+21\hat{\beta}_2$ だから
  $$\mathrm{SE}(\Delta\hat{Y})=\mathrm{SE}(\hat{\beta}_1+21\hat{\beta}_2)$$
  - $H_0:\beta_1+21\beta_2=0$ に対する F 値を用いて、$\mathrm{SE}(\Delta\hat{Y})=|\Delta\hat{Y}|/\sqrt{F}$
  - あるいは、$\mathrm{var}(\hat{\beta}_1+21\hat{\beta}_2)$ を分散と共分散に分解

# 非線形関数の重回帰推定の一般的な戦略

- ありうる非線形関係を同定する
  - 経済理論や経験から。
  - 被説明変数の変分が説明変数そのものやその他の説明変数に依存する理由があるかどうか考え直そう
- 非線形関数を特定化し OLS で係数を推定。
  - 変数間の関係の特質がわかるかもしれない
- 線形を非線形に拡張することで得るものがあるかどうかを判断
  - 「非線形だと思う」だけでは理由にならない
  - 適切かどうかの統計的判断：t 検定、F 検定
- 推定された回帰線を描いてみる
  - 当てはまっているか?
- 説明変数の変化に対する被説明変数の変化の推定

# 代表的な非線形関数

よく使われる非線形関数

- 多項式（polynomials）
- 対数（logarithms）

さしあたって説明変数がひとつのときに限定

- 説明変数が増えても拡張は簡単
- 無作為抽出標本や適切なモーメントの仮定の下で、OLS 推定量は望ましい

# 多項式関数

説明変数 $X$ の累乗（power）の次数を $r$ とおくと、多項式では

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i$$

$(X_i, Y_i)$ が i.i.d. で適当なモーメント条件を満たし、$E[u_i|X_i] = 0$ ならば、$\beta_0, \beta_1, ..., \beta_r$ は重回帰と同様に OLS で推定できる

- 線形性の検定

$$H_0 : \beta_2 = 0, \beta_3 = 0, \cdots, \beta_r = 0$$

$$\text{v.s. } H_1 : \text{少なくとも 1 つの } j \text{ に対して} \quad \beta_j \neq 0$$

を制約の数が $q = r - 1$ 個の仮説とみて F 検定すればよい

- 次数の決定
  - 柔軟さ（flexibility）と統計的正確さのトレードオフ
  - モデル化するのに必要十分な次数にすればよい……って？

# 多項式関数の次数の決定

逐次仮説検定　Sequential hypothesis testing

1. 考えられる次数 $r$ の最大値を定め、$r$ 次多項式関数を推定
2. $X^r$ の項の係数について t 検定。$\beta_r = 0$ を棄却できれば、次数は $r$ で決定
3. $\beta_r = 0$ を棄却できなければ、$r-1$ 次多項式関数を推定し、$X^{r-1}$ の項の係数について t 検定。$\beta_{r-1} = 0$ を棄却できれば、次数は $r-1$ で決定
4. $\beta_{r-1} = 0$ を棄却できなければ、$r-2$ 次多項式関数を推定し、（以下繰り返し）

最初の $r$ をどう決めるか：“spike” がなければ小さくてよく、2 から 4.

係数の解釈は単純ではないので、回帰線を描くか、$\Delta\hat{Y}$ を計算する

# 対数関数：準備

対数関数を用いると変分が変化率になるのでなにかと便利

- パーセンテージによる比較のほうがわかりやすい
- 男女間賃金格差、点数の変化、需要の価格**弾力性**$\cdots$

指数（exponential）と対数（logarithms）

- $x$ の指数関数：$e^x = \exp(x), \quad e = 2.71828...$（自然対数の底）
- $x$ の対数関数：$\ln x$

$$\ln(x + \Delta x) - \ln x = \ln\left(1 + \frac{\Delta x}{x}\right) \cong \frac{\Delta x}{x}$$

関数 $\ln(1+x)$ を 0 の周りで 1 次のテイラー展開。あるいは図を描く。

# 対数関数の 3 パターン

**Linear-log model** 説明変数のみ対数

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i, \quad \Delta Y \cong \beta_1(\Delta X/X)$$

$X$ の 1%の変化で $Y$ が $0.01\beta_1$ 変化する

**Log-linear model** 被説明変数のみ対数

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i, \quad \Delta Y/Y \cong \beta_1 \Delta X$$

$X$ の 1 単位の変化で $Y$ が $100\beta_1$%変化する

**Log-linear model** 説明変数・被説明変数ともに対数

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i, \quad \Delta Y/Y \cong \beta_1(\Delta X/X)$$

$X$ の 1%の変化で $Y$ が $100\beta_1$%変化するから、
$\beta_1$ **は** $X$ **に対する** $Y$ **の弾力性**

# 対数変換したときの被説明変数の当てはめ値

左辺が $\ln(Y_i)$ のときの $\hat{Y}_i$?

- まずは変形。両辺を指数に直して期待値をとると

$$
\begin{aligned}
E[Y_i|X_i] &= E\left[\exp\left(\beta_0 + \beta_1 X_i + u_i\right)|X_i\right] \\
&= \exp\left(\beta_0 + \beta_1 X_i\right) E\left[\exp\left(u_i\right)|X_i\right]
\end{aligned}
$$

ここで、$E[u_i|X_i] = 0$ だが、だからといって $E[e^{u_i}|X_i] = 1$ ではないから

$$
E[Y_i|X_i] \neq \exp\left(\beta_0 + \beta_1 X_i\right)
$$

– なんらかの方法で $E[e^{u_i}|X_i]$ を推定する（省略）

– $\hat{Y}_i$ を求めないで議論を展開する

# どの対数変換がよいのか

$\overline{R^2}$（自由度修正済み決定係数）

- Log-linear model と Log-linear model の比較には使ってもよい
- 通常の線形モデルと Linear-log model の比較には使ってもよい
- Log-linear model と Linear-log model の比較には使ってはいけない
  - 被説明変数が異なる場合には $\overline{R^2}$ で当てはまりを比べられない
  - $\overline{R^2}$ は被説明変数の総変動に対する当てはめ値の変動の比率

経済理論か、experts' judgement を用いる

- 賃金関数では労働経済学者は Log-linear model をよく使う
- 説明変数は少ないほうが好まれる（逐次仮説検定）
- 比較して総合的に判断。