

平成 31 年度 中級計量経済学
講義ノート 5 二項選択モデル

ここでは、被説明変数が 1 または 0 の値を取る 2 項変数である場合の計量経済モデルを取り扱う。この場合、回帰関数は、被説明変数の予測値ではなく、被説明変数が 1 をとる確率を表していると解釈できる。この節では、線形確率モデル、プロビットモデル、ロジットモデルを解説し、また最尤法による推定を紹介する。

5.1 被説明変数が 2 項変数である場合

経済学の実証分析では、被説明変数が 2 項変数である場合も多い。

例:

- 耐久消費財（家、車等）を購入するかしないか。
- （既婚女性が）働くかどうか。
- 大学に進学するかしないか。
- 喫煙するかしないか。
- 住宅ローンの審査に通るかどうか。

注：説明変数が 2 項変数である場合は講義ノート 2 で取り扱っており、特別なモデルや推定法を考える必要はない。ただ係数の解釈には注意が必要である。

被説明変数が 2 項変数であるモデルを、2 項選択モデルという。

5.2 線形確率モデル

被説明変数が 2 項変数であることを無視して線形回帰モデルの枠組みを適用すると、そのモデルはどのように解釈すればよいのだろうか。

まず、線形回帰モデルは被説明変数の条件付期待値 $E(y|\mathbf{X})$ を線形にモデル化したものであることを思い出そう。つまり、

$$E(y|\mathbf{X}) = \beta' \mathbf{X} \quad (1)$$

さて、もし y が 2 項変数なら、 $E(y|\mathbf{X}) = 1 * Pr(y = 1|\mathbf{X}) + 0 * Pr(y = 0|\mathbf{X}) = Pr(y = 1|\mathbf{X})$ である。したがって、上の式と合わせて考えると

$$E(y|\mathbf{X}) = Pr(y = 1|\mathbf{X}) = \beta' \mathbf{X} \quad (2)$$

となり、線形回帰モデルで被説明変数が 2 項変数の場合は、 $Pr(y = 1|\mathbf{X})$ という条件付確率を線形にモデル化したものと考えられることがわかる。これを線形確率モデルと呼ぶ。

- X_1 の係数 β_1 は、 X_1 を一単位変化させた時に $y = 1$ の確率がどのくらい変化するかを表わす。
- 線形確率モデルも元は上のような線形回帰モデルなので、OLS 推定やそれに基づく検定法等をそのまま適用することが可能である。
- ただし、 R^2 はそれほど意味をもたない。たとえば、被説明変数は 2 つの値しかとらないので、すべての観察点が同一直線上に並ぶことは通常ないため、 $R^2 = 1$ となることは実際にはありえない。

線形確率モデルの問題点 X の定義域に制限がない場合、モデルから予測される確率は、0 以下になったり、1 以上になったりする。そのため、線形確率モデルを使用するとモデルからの結果の解釈が困難になる場合がある。また、2 項選択モデルを使用して予測確率を計算し、それを他の統計手法を適用するときの材料に使うこともよくあり、その場合、予測確率が 0 と 1 の間に入らないと、その統計手法が使えない場合もある。

5.3 プロビット回帰とロジット回帰

線形確率モデルの問題点を解決すべく考案されたのがプロビットモデルとロジットモデルである。これらのモデルでは、モデルから予測される確率が常に 0 と 1 の間に入るが、モデルは線形でなくなる。

プロビットモデル $\Phi(\cdot)$ を標準正規分布関数として、 $\Pr(y = 1|\mathbf{X})$ を

$$\Pr(y = 1|\mathbf{X}) = \Phi(\beta'\mathbf{X}) \quad (3)$$

のようにモデル化したものをプロビットモデルという。

右辺は分布関数なので、かならず 0 と 1 の間の値をとるため、確率の予測値が $[0, 1]$ 区間をはみ出す問題は生じない。なお、どの説明変数をモデルに含めるかについては、線形回帰モデルの章で議論したように、欠落変数によるバイアスを避けることに留意して考えるとよい。

例:

- 説明変数が一つの場合。

$$\Pr(y = 1|X) = \Phi(\beta_0 + \beta_1 X). \quad (4)$$

仮に $\beta_0 = -2$ で、 $\beta_1 = 3$ であるとする。すると $X = 0.4$ の時の $Y = 1$ である確率は、

$$\Pr(y = 1|X = 0.4) = \Phi(-2 + 3 \times 0.4) = \Phi(-0.8) \approx 21.2\% \quad (5)$$

となる。

- 説明変数が二つの場合。

$$\Pr(y = 1|X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2). \quad (6)$$

仮に $\beta_0 = -1.6$ 、 $\beta_1 = 2$ で $\beta_2 = 0.5$ であるとする。説明変数の値が $X_1 = 0.4$ で $X_2 = 1$ のときの確率は

$$\Pr(y = 1|X_1 = 0.4, X_2 = 1) = \Phi(-0.3) \approx 38\% \quad (7)$$

である。

X を変化させたときの影響 プロビットモデルは非線形モデルである。そのため、 X を変化させたときの影響を X の係数の値だけでとらえることができない。以下のように、変化前と変化後の予測確率の差を計算する必要がある。 X_1 を ΔX_1 だけ変化させたときの影響は

$$\Pr(y = 1|X_1 + \Delta X_1, X_2, \dots, X_k) - \Pr(y = 1|X_1, X_2, \dots, X_k) \quad (8)$$

$$= \Phi(\beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2 + \dots + \beta_k X_k) - \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) \quad (9)$$

であり、これより簡単にはできない。

限界効果は、合成関数の微分を用いて次のように計算できる。

$$\frac{\partial \Phi(\beta' \mathbf{X})}{\partial X_k} = \phi(\beta' \mathbf{X}) \beta_k. \quad (10)$$

これは \mathbf{X} が異なれば異なる値をとる。そこで、論文などでは、代表的な値として \mathbf{X} の平均あるいは中央値での限界効果を計算して掲載することが多い。

ロジットモデル プロビットモデルでは、標準正規分布関数を使用した。そのかわりに標準ロジスティック分布関数 $\Lambda(\cdot)$ を使用した条件付確率モデルをロジットモデルという。そのモデルでは、 $y = 1$ の条件付き確率は

$$\Pr(y = 1 | X_1, \dots, X_k) = \Lambda(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) \quad (11)$$

$$\equiv \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}} \quad (12)$$

となる。

- 歴史的には、ロジットモデルは推定時の計算がプロビットモデルに比べて早いためによく使われてきた。ただ、現代の計算機では気にするほどの違いではない。
- ロジットモデルとプロビットモデルは非常に似通った推定結果を与えることが多い。
- モデルをもっと複雑にしていくと二つのモデルの利点には違いが出てくる。具体的には、2項選択ではなく、多項選択モデルを考えると、プロビットを拡張すると、色々な選択肢間の相関を簡単に表現できるが、推定量の計算が難しくなる。一方、ロジットモデルを拡張した場合は、選択肢間の相関に制約を加えないとうまくモデル化できないが、計算は比較的容易になる。

線形確率モデルとプロビットやロジットモデルとの比較 どのモデルを使うべきかについては、特に決まった答えがあるわけではない。どのモデルがもっともデータと整合的かを調べる統計手法は存在するが、それほど頻繁には使われていない。

ただ、Angrist and Pischke (2008) の意見は一考の価値がある。線形確率モデルを使って推定をした場合、たとえ真の条件付期待値が線形でなかったとしても、なんらかの意味のあるパラメータを推定することができる。たとえば、回帰変数が2項変数の場合、OLS 推定量は平均処置効果 ($E(y|X=1) - E(y|X=0)$) に収束する。しかし、プロビットやロジットの場合、もしモデルが間違っている場合には推定結果の解釈が難しくなる。したがって、線形確率モデルを使用すべきということである。

ただし、さきに見たとおり、線形確率モデルでは、確率としては不適切な値を確率の予測値をしてしまう可能性がある。

5.4 プロビットモデルの推定

プロビットモデルは係数に関して非線形なモデルである。したがって、OLS で推定することはできない。ここでは最尤法での推定を考える。

なお、ロジットモデルの場合も同じように推定することができる。

最尤推定法 (MLE) 得られたデータが観察される確率 (密度) を未知パラメータの関数とみたものを尤度関数という。尤度関数を最大にするようにパラメータの値を決める推定法を最尤推定といい、その推定量を最尤推定量という。

アイデア:

- 実際に観察されたデータを最も高い確率で生成するような係数の値を、推定値として選ぶ。

最初に、説明変数がない簡単な場合を考えよう。 $p = \Pr(y = 1)$ とする。すると、 $y_i, i = 1, \dots, n$ という標本が観察される確率は、

$$\prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} \quad (13)$$

である。この確率をパラメータ p に関する関数とみなしたものが尤度関数である。これを p に関して最大にすればよいが、計算しにくい。そこで、尤度関数の自然対数をとったものを考える。これを対数尤度関数といい、この場合は

$$L(p) = \sum_{i=1}^n \{y_i \log p + (1-y_i) \log(1-p)\} \quad (14)$$

である。対数変換は単調なので、尤度関数を最大にするパラメータの値は対数尤度関数を最大にする値と同じである。尤度関数は積の形になっているが、対数尤度は和の形になっているので、最大化のための微分が簡単になる。そのため、通常は対数尤度を用いて最大化を行う。 p の最尤推定量は $L(p)$ の最大値をもたらす p の値であり、この例では、 $\hat{p} = \bar{y}$ であることが簡単に確かめられる。

次に、プロビットモデルを考える。プロビットモデルでは、 $\Pr(y_i = 1 | \mathbf{X}_i) = \Phi(\mathbf{X}_i' \beta)$ である。このとき、 X_i を条件とする対数尤度関数は

$$L(\beta) = \sum_{i=1}^n \{y_i \log \Phi(\mathbf{X}_i' \beta) + (1-y_i) \log(1 - \Phi(\mathbf{X}_i' \beta))\} \quad (15)$$

である。なお、 X_i の分布が、 β に依存しないのなら、これを尤度として考えても問題はなく、通常そのように仮定する (X_i は弱外生であるという)。そして、 $L(\beta)$ を最大化することによって、 β の最尤推定量を得る。ただし、プロビットモデルの最尤推定量を明示的な式で書くことができない。したがって、数値計算によって対数尤度関数を最大化する必要があるが、多くの統計ソフトで自動的に計算してくれる。

5.5 最尤推定量の漸近的性質

最尤推定量は一般に明示的な式で書くことができず、その漸近的性質を直接証明することができない。

ここでは、どのようにして最尤推定量の漸近的性質を求めるのかを、かなりいい加減にみていく。数学的に厳密な証明は、Hayashi (2000) をはじめとする上級の計量経済学の教科書を参照のこと。

まず、最尤推定量の一致性は、尤度関数の極限を考え、それが真のパラメータの値で最大化されることを示し、最大値をもたらす統計量はその関数の極限の最大値をもたらす値に収束することを証明することで示す。つまり、

$$\frac{1}{n} L(\beta) \rightarrow_p E(y_i \log \Phi(\mathbf{X}_i' \beta) + (1-y_i) \log(1 - \Phi(\mathbf{X}_i' \beta))) \quad (16)$$

ということを示し、 $\beta_0 = \arg \max E(y_i \log \Phi(\mathbf{X}'_i \beta) + (1 - y_i) \log(1 - \Phi(\mathbf{X}'_i \beta)))$ ということを証明する。すると、

$$\hat{\beta} = \arg \max \frac{1}{n} L(\beta) \rightarrow_p \arg \max E(y_i \log \Phi(\mathbf{X}'_i \beta) + (1 - y_i) \log(1 - \Phi(\mathbf{X}'_i \beta))) = \beta_0 \quad (17)$$

ということが証明できる。

次に最尤推定量の漸近分布の導出の仕方を紹介する。基本的な方針は、テーラー展開によって最尤推定量を線形化し、そして中心極限定理や大数の法則を使用するというものである。

まず、 $g(\cdot)$ を $L(\cdot)$ の一次微分関数であるとする。すると最尤推定量は次の式を満たす。

$$g(\hat{\beta}) = 0. \quad (18)$$

この式を真のパラメーターの値の周りでテーラー展開をすると、

$$g(\beta_0) + H(\beta_0)(\hat{\beta} - \beta_0) + small = 0 \quad (19)$$

となる。ここで $H(\cdot)$ は $L(\cdot)$ の2回微分関数である。この方程式を書き直すと、

$$\sqrt{n}(\hat{\beta} - \beta_0) = - \left(\frac{1}{n} H(\beta_0) \right)^{-1} \frac{1}{\sqrt{n}} g(\beta_0) + small \quad (20)$$

となる。ここで、 $H(\beta_0)$ と $g(\beta_0)$ は、i.i.d. の確率変数の和で書くことができる。 $H(\beta) = \sum_{i=1}^n H_i(\beta)$ かつ $g(\beta) = \sum_{i=1}^n g_i(\beta)$ と定義する。すると

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow_d N \left(\mathbf{0}, (E(H_i(\beta_0)))^{-1} E(g_i(\beta_0)g_i(\beta_0)') (E(H_i(\beta_0)))^{-1} \right) \quad (21)$$

となり、漸近分布が導出できる。

ところで、次の情報等式と呼ばれる式が成立する。

$$-E(H_i(\beta_0)) = E(g_i(\beta_0)g_i(\beta_0)'). \quad (22)$$

なお、この等式は、真の値 β_0 の元でしか成立しない。情報等式を使うと、

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow_d N \left(\mathbf{0}, - (E(H_i(\beta_0)))^{-1} \right) = N \left(\mathbf{0}, (E(g_i(\beta_0)g_i(\beta_0)'))^{-1} \right) \quad (23)$$

となる。

$\hat{\beta}$ の漸近分散の推定は二つの方法で行うことができる。

- $-\left(\frac{1}{n} H(\hat{\beta})\right)^{-1}$.
- $\left(\frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta})g_i(\hat{\beta})'\right)^{-1}$. これは、BHHH (Berndt, Hall, Hall and Hausman) 推定量と呼ばれる。 H の式を導出するのが難しい場合にはこの推定量は有用であるが、 H が明示的に出るとなると、避けたほうがよい。

5.6 被説明変数が2項変数である場合の当てはまりの良さの指標

先にみたとおり、被説明変数が2項変数の場合には、 R^2 や \bar{R}^2 は適切な指標ではない。代わりに次のような指標が用いられる。

- 当たり率: もし、予測確率が0.5より大きければ $y = 1$ を予測値とし、予測確率が0.5より小さければ $y = 0$ と予測値としよう。そのとき、予測が当たった割合を考える。それは、 $\hat{p}_i = \Phi(\mathbf{X}'_i \hat{\beta})$ として、次のように表すことができる。

$$R_{hitrate} = \frac{1}{n} \sum_{i=1}^n (1 - |y_i - \mathbf{1}(\hat{p}_i > 0.5)|) \quad (24)$$

- ただ、当たり率はその解釈が難しい可能性がある。特に、ほとんどの観測値で、 $y_i = 0$ となっている場合、モデルに関わらず $y = 0$ を予測することで高い当たり率を得ることができる。かわりに次の指標を考える。 \hat{p} を $y_i = 1$ である観測値の割合であるとする。 $wr_0 = \min(\hat{p}, 1 - \hat{p})$ とする。そして次の指標を考える。

$$R_p^2 = 1 - \frac{1 - R_{hitrate}}{wr_0} \quad (25)$$

- 疑似 R^2 : 切片を除くすべての係数が0であるという制約のもとで最大化した対数尤度関数の値を L_0 とする。それは、 $n_1 = \sum_{i=1}^n y_i$ として

$$L_0 = n_1 \log(n_1/n) + (n - n_1) \log(1 - n_1/n), \quad (26)$$

と書ける。これを用いて、疑似 R^2 は

$$pseudoR^2 = 1 - \frac{1}{1 + 2(L(\hat{\beta}) - L_0)/n} \quad (27)$$

と定義される。一般には $L(\hat{\beta}) > L_0$ であるが、 X が説明力を持たないときは、ほぼ $L(\hat{\beta}) = L_0$ となる。

- McFadden の R^2 :

$$McFaddenR^2 = 1 - \frac{L(\hat{\beta})}{L_0}. \quad (28)$$

$0 > L(\hat{\beta}) > L_0 >$ であることに注意すると、この指標は1と0の間の値をとることがわかる。なお、この指標も疑似 R^2 と呼ばれることがある。統計ソフトによって疑似 R^2 の定義が異なる場合があるので、注意が必要である。

5.7 経済学的モデルからプロビットモデルを導出する

確率的効用モデル 経済学的な意思決定モデルから2項選択モデルを導出することも可能である。例えば、大学に進学するかどうかの意思決定問題を考えよう。個人 i が大学に行くとき $y_i = 1$ 、行かないとき $y_i = 0$ とする。

- 個人 i が大学に行くとき、次の効用を得るとする。

$$u_{1i} = \mathbf{X}_i' \beta_1 + \epsilon_{1i}. \quad (29)$$

ここで、 \mathbf{X}_i は大学に行くことによる効用に影響を与える観測できる変数とし、 ϵ_{1i} は効用に影響を与える他の要因を表すとする。

- 大学に行かないときに得られる効用も次のように決まるとしよう。

$$u_{0i} = \mathbf{X}_i' \beta_0 + \epsilon_{0i}. \quad (30)$$

- ϵ_{1i} と ϵ_{0i} は、2次元正規分布に従っているとする。
- 人々は効用最大化によって行動しているとする ($u_{0i} < u_{1i}$ なら大学に行く)。

すると、 $y_i = 1$ である (大学に進学する) 確率は、

$$\Pr(y_i = 1 | \mathbf{X}_i) = \Pr(u_{1i} > u_{0i} | \mathbf{X}_i) = \Pr(\epsilon_{0i} - \epsilon_{1i} < \mathbf{X}_i' (\beta_1 - \beta_0)) = \Phi(\mathbf{X}_i' \beta^*) \quad (31)$$

である。ここで、 $\beta^* = (\beta_1 - \beta_0) / \sqrt{\text{var}(\epsilon_{0i} - \epsilon_{1i})}$ である。

上のような確率的効用モデルに従って人々が大学に行くかどうかの選択を行っているなら、データ生成過程はプロビットモデルで記述できることになる。この結果は、プロビットモデルの係数を解釈する際に役に立つ。例えば、

- 係数の大きさそのものは、あまり意味がない。(効用はその大きさそのものには意味がないので。)
- 係数は、ある変数が二つの選択肢それぞれから来る効用に与える影響の「差」として解釈できる。

確率的効用モデルによる2項選択モデルの基礎付けは、近年広く使われている構造推定におけるモデル構築の基礎をなすものであり、非常に重要である。

潜在変数モデル 潜在変数モデルにおいても、データ生成過程をプロビットモデルで記述できる。 $u_i \sim N(0, 1)$ として

$$y_i^* = \mathbf{X}_i' \beta + u_i \quad (32)$$

と定義する。 y_i^* それ自体は観測できないが、 $y_i^* > 0$ であるかどうかは観測できるとする。 $y_i = 1\{y_i^* > 0\}$ と定義すると、データ生成過程はプロビットモデルで記述できる。

- 潜在変数モデルは、効用の差それ自体をモデル化したものとも考えることもできる。
- 潜在変数モデルが適切である例としては、所得は観測できないが、貧困状態にあるかどうかはわかる場合などである。

5.8 分散不均一性

プロビットモデルにおいて分散不均一性を考える際には、2つの可能性がある。まず、プロビットモデルを回帰の枠組みで捉えた時、以下のようにそれ自体が分散不均一なモデルであることがわかる。被説明変数の平均と分散は、

$$E(y_i|\mathbf{X}_i) = \Phi(\mathbf{X}'_i\beta), \quad (33)$$

$$\text{var}(y_i|\mathbf{X}_i) = \Phi(\mathbf{X}'_i\beta)(1 - \Phi(\mathbf{X}'_i\beta)) \quad (34)$$

で、分散が \mathbf{X} の値に依存し、特に特別な形の分散不均一になっていることがわかる。これは2項分布の期待値と分散の関係から来ている。

また、潜在変数モデルにおける分散不均一性も考えることができる。次の潜在変数モデルを考える。

$$y_i^* = \mathbf{X}'_i\beta + u_i. \quad (35)$$

このモデルがプロビットモデルをもたらすためには、誤差項 u_i は分散均一である必要がある。 u_i が分散不均一であるような拡張は可能である。しかし、そうした拡張は色々な理論上や計算上の注意が必要なことが知られており、この講義では取り扱わない。

- もし u_i が分散不均一なら、通常のプロビットモデルを最尤推定しても係数の一致推定量は得られない。
- もし u_i が分散不均一なら、 y_i の条件付き期待値は $\Phi(\mathbf{X}'_i\beta)$ ではなくなる。

なお、線形確率モデルであれば、分散不均一性は通常 of 線形回帰モデルと同様に取り扱うことができる。なお、常に分散不均一になるので、頑健な漸近分散推定量を使用する必要がある。