平成 **28** 年度　上級計量経済学

講義ノート **6:** 被説明変数のとりうる値に制約のあるモデル **(Limited dependent variable models)**

このノートでは、質的選択モデルや切断回帰モデル、打ち切り回帰モデルなどといった、被説明変数のとりうる値に制約のかかっているモデルの紹介と、その解説を行う。

従属変数が、家や車などの耐久消費財を保有しているか否か、といった選択データの分析には質的選択モデル（qualitative response model）が用いられる。その最も基本的なモデルが 2 値選択モデル（binary choice model）である。

また、被説明変数が、ある閾値を越えるとデータに出てこなくなる場合は、切断回帰モデル (truncated regression model)、被説明変数のみが観測できなくなる場合は打ち切り回帰モデル (censored regression model) が用いられる。

6.1　2 値選択モデル

y_i を 0 か 1 の値を取る被説明変数、X_i を k 次元の説明変数、F をある分布関数として、

$$P(y_i = 1|X_i) = F(X_i'\beta) \tag{1}$$

を 2 値選択モデルという。(1) は、以下のような状況から導出される。

1. 潜在変数モデル (Latent variable model)

   例えば、X を個人の属性として、家を保有している場合は $y = 1$、保有していない場合は $y = 0$ とする。個人 i は $y_i^* = X_i'\beta + \epsilon_i$ の値が正か負かによって意思決定を行うが、y_i^* 自体は観測されない変数（latent variable）である。y^* が正の場合は $y = 1$、負の場合は $y = 0$ となるものとする。つまり、

$$\begin{aligned} y_i^* &= X_i'\beta + \epsilon_i && (2)\\ y_i &= I(y_i^* \geq 0) && (3) \end{aligned}$$

   としてデータは生成される。また、X_i と ϵ_i は独立とし、$-\epsilon_i$ の分布関数を F とする。そのとき、

$$\begin{aligned} P(y_i = 1|X_i) &= P(y_i^* \geq 0|X_i) && (4)\\ &= P(X_i'\beta + \epsilon_i \geq 0|X_i) && (5)\\ &= P(-\epsilon_i \leq X_i'\beta|X_i) && (6)\\ &= F(X_i'\beta) && (7) \end{aligned}$$

   となる。

2. 確率的効用モデル

   例えば、個人 i が通勤に車を使う $(y_i = 1)$ か、公共交通機関を使う $(y_i = 0)$ かを決める問題を考える。前者の場合の効用は $u_{1i} = z_{1i}'\beta + \epsilon_{1i}$、後者の場合の効用は $u_{0i} = z_{0i}'\beta + \epsilon_{0i}$ とする。z は説明変数で、たとえば所要時間、費用等である。また、$\epsilon_{0i} - \epsilon_{1i}$ の分布関数を F とし、$X_i = z_{1i} - z_{0i}$ とする。効用最大化を仮定すると、個人 i の意思決定は次のように表される。

$$\begin{aligned} P(y_i = 1|X_i) &= P(u_{1i} \geq u_{0i}|X_i) && (8)\\ &= P\{(z_{1i}'\beta + \epsilon_{1i}) - (z_{0i}'\beta + \epsilon_{0i}) \geq 0|X_i\} && (9)\\ &= P(\epsilon_{0i} - \epsilon_{1i} \leq (z_{1i} - z_{0i})'\beta|X_i) && (10)\\ &= F(X_i'\beta) && (11) \end{aligned}$$

実証上よく用いられる特定化は

- 線形確率モデル：$F(x) = x$。
  この場合は OLS 推定することが可能である。ただし、線形確率モデルでは x が負値や 1 より大きい値はとれないので、注意して適用する必要がある。

- プロビット (probit) モデル：
$$F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{12}$$
  (正規分布)

- ロジット (logit) モデル：
$$F(x) = \Lambda(x) = \frac{\exp(x)}{1+\exp(x)} \tag{13}$$
  (ロジスティック分布)

プロビットモデルやロジットモデルは、y_i を X_i に回帰する OLS では推定できない。なぜなら、単純化のために説明変数 X はスカラー、定数項を含めるものとして

$$E(\beta_{OLS}|X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})E(y_i|X_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})F(X_i'\beta)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \neq \beta \tag{14}$$

だからである。β は ML 推定することができる。

6.2 2 値選択モデルの最尤推定

2 値選択モデルを最尤推定を議論するために、まずは尤度関数を導出する。(y_i, X_i), $i = 1, \dots, n$ を無作為標本とする。各個人の確率関数は

$$\begin{aligned} P(y_i = 1|X_i) &= F(X_i'\beta) & (15) \\ P(y_i = 0|X_i) &= 1 - F(X_i'\beta) & (16) \end{aligned}$$

なので、

$$P(y_i = y|X_i) = F(X_i'\beta)^y[1 - F(X_i'\beta)]^{1-y} \ \ (y = 1, 0) \tag{17}$$

と書き換えることができる。したがって、尤度関数は、

$$L(\beta) = \prod_{i=1}^{n} F(X_i'\beta)^{y_i}[1 - F(X_i'\beta)]^{1-y_i} \tag{18}$$

であり、対数尤度関数は

$$\log L(\beta) = \sum_{i=1}^{n} y_i \log F(X_i'\beta) + (1 - y_i)\log[1 - F(X_i'\beta)] \tag{19}$$

となる。また、パラメータ空間を Θ として、最尤推定量は

$$\hat{\beta} = \arg\max_{\beta\in\Theta} \log L(\beta) \tag{20}$$

で定義される。

この推定量は、$W_i = (y_i, X_i)$、

$$m(W_i; \beta) = y_i \log F(X_i'\beta) + (1 - y_i) \log[1 - F(X_i'\beta)] \tag{21}$$

とすると、M 推定量の特殊ケースになり、極値推定量の理論を用いて、漸近的性質を調べることができる。以下では、条件付き最尤推定量に限って漸近理論を述べる。条件付き最尤推定量というのは、X_i を条件づけた尤度関数を考えており、X_i の密度も考慮に入れたデータ全体の分布を元にした尤度関数を考えている訳ではないからである。

6.3 一般的な最尤推定量の漸近理論

一般的な (条件付き) 最尤推定量の漸近理論を述べよう。以下では、(y_i, X_i)、$i = 1, \ldots, n$ は i.i.d. であると仮定する。X_i を条件とする y_i の条件付き密度（または確率）は $f(y_i|X_i; \theta)$ であるとする。そのとき、θ の最尤推定量は、

$$\hat{\theta} = \arg\max_{\theta} \log L(\theta) = \sum_{i=1}^{n} \log f(y_i|X_i; \theta) \tag{22}$$

である。

命題 1 (条件付き最尤推定量の一致性（母数空間がコンパクトな場合）)**.** 以下の仮定をおく。
(i) パラメータ空間 Θ は k 次元ユークリッド空間のコンパクトな集合である。
(ii) $f(y_i|X_i; \theta)$ はすべての (y_i, X_i) に対して θ について連続である。
*(iii) (*識別*)* 全ての θ_0 でない $\theta \in \Theta$ について、$P[f(y_i|X_i; \theta) \neq f(y_i|X_i; \theta_0)] > 0$
*(iv) (*押さえ込み*)* $E[\sup_{\theta \in \Theta} |\log f(y_i|X_i; \theta)|] < \infty$

そのとき、

$$\hat{\theta} \to_p \theta_0 \tag{23}$$

(iii) は若干回りくどい表現に思われるが、言いかえると、$\theta = \theta_0$ のときのみ $f(y_i|X_i; \theta) = f(y_i|X_i; \theta_0)$ ということである。次に、パラメータ空間のコンパクト性を仮定しない結果を示す。

命題 2. 条件付き（疑似）最尤推定量の一致性（コンパクト性を仮定しない場合）
以下の仮定をおく。
(i)' パラメータ空間 Θ は k 次元ユークリッド空間の凸集合で、真の値 θ_0 は Θ の内点である。
(ii)' $\log f(y_i|X_i; \theta)$ はすべての (y_i, X_i) に対して θ について凹 *(concave)* である。
*(iii) (*識別*)* $P[f(y_i|X_i; \theta) \neq f(y_i|X_i; \theta_0)] > 0$, $\forall\theta$, $\theta \neq \theta_0$, $\theta \in \Theta$
*(iv)' (*押さえ込み*)* $E[|\log f(y_i|X_i; \theta)|] < \infty$, $\forall\theta \in \Theta$

そのとき、

$$\hat{\theta} \to_p \theta_0 \tag{24}$$

情報不等式　上の二つの定理の識別条件が、極値推定量の識別条件になっていることは、情報不等式と呼ばれる不等式によって、示すことができる。密度関数 $f(x, \theta)$ によって最尤推定する場合を考える。また真の分布は $f(x, \theta_0)$ であるとする。このとき Jensen の不等式によって、

$$E\left\{\log\left[\frac{f(x, \theta)}{f(x, \theta_0)}\right]\right\} \leq \log E\left[\frac{f(x, \theta)}{f(x, \theta_0)}\right] \tag{25}$$

が成り立つ。しかし、

$$E\left[\frac{f(x,\theta)}{f(x,\theta_0)}\right]=\int\frac{f(x,\theta)}{f(x,\theta_0)}f(x,\theta_0)dx=\int f(x,\theta)dx=1 \tag{26}$$

が成り立つ。よって、

$$E\left\{\log\left[\frac{f(x,\theta)}{f(x,\theta_0)}\right]\right\}\leq 0 \tag{27}$$

であり、ここから、

$$E[\log f(x,\theta)]\leq E[\log f(x,\theta_0)] \tag{28}$$

となる。つまり、対数尤度の極限の関数は、真値 θ_0 で最大値をとる。Jensen の不等式の条件から、確率 1 で $f(x,\theta)$ と $f(x,\theta_0)$ が一致しない限り、不等式は強い不等式であることがわかる。したがって、極値推定量の識別条件は満たされることになる。

次に、漸近正規性を述べる。命題 1 または 2 の条件が成立し、従って $\hat{\theta}\to_p\theta_0$ であるとする。

命題 3 (条件付き最尤推定量の漸近正規性)**.** 以下を仮定する。

(i) θ_0 は Θ の内点である。(命題 *2* では仮定済み)

(ii) $f(y_i|X_i;\theta)$ はすべての (y_i,X_i) に対して θ について 2 回連続微分可能である。

(iii) $s(y_i,X_i;\theta)=\partial\log f(y_i|X_i;\theta)/(\partial\theta)$、$H(y_i,X_i;\theta)=\partial^2\log f(y_i|X_i;\theta)/(\partial\theta\partial\theta')$ として、

$$E[s(y_i,X_i;\theta_0)]=0,\ \ -E[H(y_i,X_i;\theta_0)]=E[s(y_i,X_i;\theta_0)s(y_i,X_i;\theta_0)'] \tag{29}$$

が成り立つ。

*(iv) (*ヘシアン行列の近傍での押さえ込み条件*)* θ_0 の近傍 $\mathcal{N}$ に対して、

$$E[\sup_{\theta\in N}||H(y_i,X_i;\theta)||]<\infty \tag{30}$$

であり、従って任意の $\bar{\theta}\to_p\theta_0$ に対して

$$\frac{1}{n}\sum_{i=1}^{n}H(y_i,X_i;\bar{\theta})\to_p E[H(y_i,X_i;\theta_0)] \tag{31}$$

(v) $I(\theta_0)=-E[H(y_i,X_i;\theta_0)]$ は非特異である。

そのとき、

$$\sqrt{n}(\hat{\theta}-\theta_0)\to_d N(0,I(\theta_0)^{-1}) \tag{32}$$

である。

スコア関数 $s(y_t,x_t,\theta_0)$ の期待値が 0 となることは、$w=(y,x)$ として $a(w;\theta)=f(w,\theta)$ について、以下の微分と積分の入れ替えが成立すれば成立する。

$$\frac{\partial}{\partial\theta}\int a(w,\theta)dw=\int\frac{\partial}{\partial\theta}a(w,\theta)dw \tag{33}$$

通常の計量経済モデルではこれは成立するが、そのための十分条件は Newey and McFadden (1994, Lemma 3.6, p.2152) を参照。$f(x,\theta)$ を密度関数とする場合を考えると、

$$\int f(x,\theta)dx=1 \tag{34}$$

であるので、両辺を微分して、

$$\frac{\partial}{\partial\theta}\int f(x,\theta)dx = 0 \tag{35}$$

であるが、積分と微分を入れ替えると、

$$0 = \int \frac{\partial}{\partial\theta} f(x,\theta)dx = \int \frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)dx = \int s(x,\theta)f(x,\theta)dx \tag{36}$$

となるからである。

情報等式　(iii) の二つめの不等式は情報等式と呼ばれる。これは、$w=(y,X)$ として $a(w;\theta) = [\partial \log f(w,\theta)/(\partial\theta)]f(w;\theta)$ について、微分と積分の入れ替え

$$\frac{\partial}{\partial\theta}\int a(w,\theta)dw = \int \frac{\partial}{\partial\theta} a(w,\theta)dw \tag{37}$$

ができれば、情報等式は成り立つ。この入れ替えの十分条件は Newey and McFadden (1994, Lemma 3.6, p.2152) を参照。証明は先ほどと同じように、$f(x,\theta)$ を密度関数とする場合を考えると、

$$0 = \int s(x,\theta)f(x,\theta)dx \tag{38}$$

が成り立ち、これをさらに微分すると、

$$0 = \int H(x,\theta)f(x,\theta)dx + \int s(x,\theta)s(x,\theta)dx \tag{39}$$

となるからである。

6.4 ロジットモデルとプロビットモデルの MLE の漸近的性質

以下に、命題 2 に基づいてロジットモデル、プロビットモデルの場合について (20) の解の一致性を示す。なお、微分可能性等の条件のもとでは、$f(x) = dF(x)/dx$ とすると、MLE の一階の条件は

$$\frac{\partial \log L(\hat\beta)}{\partial\beta} = \sum_{i=1}^{n} \frac{y_i - F(X_i'\hat\beta)}{F(X_i'\hat\beta)[1-F(X_i'\hat\beta)]} f(X_i'\hat\beta)X_i = 0 \tag{40}$$

となる。また、2 階の条件は

$$\begin{aligned}\frac{\partial^2 \log L(\hat\beta)}{\partial\beta\partial\beta'} &= -\sum_{i=1}^{n}\left\{\frac{y_i - F(X_i'\hat\beta)}{F(X_i'\hat\beta)[1-F(X_i'\hat\beta)]}\right\}^2 f(X_i'\hat\beta)^2 X_i X_i' \\ &\quad + \sum_{i=1}^{n}\frac{y_i - F(X_i'\hat\beta)}{F(X_i'\hat\beta)[1-F(X_i'\hat\beta)]} f'(X_i'\hat\beta)X_iX_i' \\ &< 0 \end{aligned} \tag{41}$$

である。

定理 1 (ロジットモデルとプロビットモデルの MLE の一致性). $\Theta = R^k$ で $E||X_iX_i'|| < \infty$, $E(X_iX_i')$ が非特異行列であるとき、

$$\hat\beta \to_p \beta_0 \tag{42}$$

である。

（証明） 命題 2 の各条件が成立していることを示せばよい。(i)' は明らかである。

(ii)' について、ロジットモデルの場合のみ示す（プロビットの場合は若干厄介であるが、Amemiya (1985, p.273-274; p273 最終式に誤植の可能性あり) を参照)。簡単な計算から

$$f(u) = \Lambda'(u) = \Lambda(u)[1 - \Lambda(u)] \tag{43}$$

$$f'(u) = \Lambda''(u) = [1 - 2\Lambda(u)]\Lambda(u)[1 - \Lambda(u)] \tag{44}$$

であることがわかる。したがって、(41) の等式部分に $F = \Lambda$ を代入して 2 次微分を計算すると、$y_i(1 - y_i) = 0$ を用いて

$$\frac{\partial^2 \log L(\beta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{n} \Lambda(X_i'\beta)[1 - \Lambda(X_i'\beta)]X_iX_i' \tag{45}$$

であることがわかる。$0 < \Lambda(X_i'\beta) < 1$ また $E(X_iX_i') > 0$ なので、(ii)' が成立する。

(iii) 仮定より $E(X_iX_i')$ は正値定符号であるから、$E(X_i'\beta - X_i'\beta_0)^2 = (\beta-\beta_0)'E(X_iX_i')(\beta-\beta_0)$ より、$\beta \neq \beta_0$ なら $E(X_i'\beta - X_i'\beta_0)^2 \neq 0$ である。従って、$\beta \neq \beta_0$ のとき $P(X_i'\beta \neq X_i'\beta_0) > 0$ でなければならない。すると、正規分布、ロジット分布の分布関数は狭義の増加関数なので、$P[F(X_i'\beta) \neq F(X_i'\beta_0)] > 0$ が成立する。さて、

$$f(y_i|X_i;\beta) \;\; = \;\; F(X_i'\beta)^{y_i}[1 - F(X_i'\beta)]^{1-y_i} \tag{46}$$

なので、

$$P[f(y_i|X_i;\beta) = f(y_i|X_i;\beta_0)] \tag{47}$$
$$= P\{F(X_i'\beta)^{y_i}[1 - F(X_i'\beta)]^{1-y_i} = F(X_i'\beta_0)^{y_i}[1 - F(X_i'\beta_0)]^{1-y_i}\} \tag{48}$$
$$= P\{F(X_i'\beta)^{y_i}[1 - F(X_i'\beta)]^{1-y_i} = F(X_i'\beta_0)^{y_i}[1 - F(X_i'\beta_0)]^{1-y_i}|y_i = 1\}P(y_i = 1) \tag{49}$$
$$+ P\{F(X_i'\beta)^{y_i}[1 - F(X_i'\beta)]^{1-y_i} = F(X_i'\beta_0)^{y_i}[1 - F(X_i'\beta_0)]^{1-y_i}|y_i = 0\}P(y_i = 0) \tag{50}$$
$$= P[F(X_i'\beta) = F(X_i'\beta_0)|y_i = 1]P(y_i = 1) \tag{51}$$
$$+ P[F(X_i'\beta) = F(X_i'\beta_0)|y_i = 0]P(y_i = 0) \tag{52}$$
$$= P[F(X_i'\beta) = F(X_i'\beta_0)] \tag{53}$$

となる。両辺をそれぞれ 1 から引くと

$$P[f(y_i|X_i;\theta) \neq f(y_i|X_i;\theta_0)] = P[F(X_i'\beta) \neq F(X_i'\beta_0)] > 0 \tag{54}$$

を得る。なお、ここで $f(y_i|X_i;\beta)$ は密度ではなく確率である。

(iv)' $\log f(y_i|X_i;\beta) = y_i \log F(X_i'\beta) + (1 - y_i)\log[1 - F(X_i'\beta)]$ なので、$y_i = 0, 1$、$F(-u) = 1 - F(u)$ に注意して、

$$E[|\log f(y_i|X_i;\beta)|] \leq E[|\log F(X_i'\beta)|] + E[|\log F(-X_i'\beta)|] \tag{55}$$

である。まず、プロビットモデル（$F = \Phi$）のとき、$|\log\Phi(u)| \leq |\log\Phi(0)| + |u| + u^2$ (Appendix を参照) であり、仮定から $E(X_iX_i')$ は有界なので、

$$E[|\log\Phi(X_i'\beta)|] + E[|\log\Phi(-X_i'\beta)|] \leq 2|\log\Phi(0)| + 2E|X_i'\beta| + 2E(X_i'\beta)^2 < \infty \tag{56}$$

となる。次にロジットモデル（$F = \Lambda$）の場合を考える。$u \geq 0$ なら、$\Lambda(0) \leq \Lambda(u) < 1$ なので、

$$|\log\Lambda(u)| \leq |\log\Lambda(0)| \leq |u| + |\log\Lambda(0)| \tag{57}$$

である。$\exp(\cdot)$ と $\log(\cdot)$ は単調増加関数なので、$u<0$ なら

$$0 > \log\left[\frac{1}{1+\exp(u)}\right] > \log\left[\frac{1}{1+\exp(0)}\right] = \log\Lambda(0) \tag{58}$$

である。よって

$$|\log\Lambda(u)| = \left|u + \log\left[\frac{1}{1+\exp(u)}\right]\right| < |u+\log\Lambda(0)| < |u| + |\log\Lambda(0)| \tag{59}$$

なので、任意の u について $|\log\Lambda(u)| \leq |u| + |\log\Lambda(0)|$ が成り立つ。よって、$E|\beta' X_i| < \infty$ なら押さえ込み条件 (dominance condition) が成り立つが、それは $E(X_i X_i')$ の有界性から成立する。(証明終)

定理 2. 漸近正規性

$\Theta = R^k$ で $E(X_i X_i')$ が非特異行列で $E(\|X_i X_i'\|) < \infty$ であるとき、

$$\sqrt{n}(\hat\beta - \beta_0) \xrightarrow{d} N(0, I(\beta_0)^{-1}) \tag{60}$$

である。

(証明) 命題 3 の (i)-(v) を示せばよい。(i) は $\Theta = R^k$ として、成立する。(ii) は明らかに成立する。

(iii) スコアとヘシアンは $F = \Lambda, \Phi$ に対して

$$\begin{aligned} s(y_i, X_i; \beta_0) &= \frac{\partial \log f(y_i|X_i;\beta_0)}{\partial\beta} && (61)\\ &= \frac{y_i - F(X_i'\beta_0)}{F(X_i'\beta_0)[1-F(X_i'\beta_0)]} f(X_i'\beta_0) X_i && (62)\end{aligned}$$

$$\begin{aligned} H(y_i, X_i; \beta_0) &= \frac{\partial^2 \log f(y_i|X_i;\beta_0)}{\partial\beta\partial\beta'} && (63)\\ &= -\left\{\frac{y_i - F(X_i'\beta_0)}{F(X_i'\beta_0)[1-F(X_i'\beta_0)]}\right\}^2 f(X_i'\beta_0)^2 X_i X_i' && (64)\\ &\quad + \frac{y_i - F(X_i'\beta_0)}{F(X_i'\beta_0)[1-F(X_i'\beta_0)]} f'(X_i'\beta_0) X_i X_i' && (65)\end{aligned}$$

である。$E(y_i|X_i) = E(y_i^2|X_i) = F(X_i'\beta_0)$ なので、$E[s(y_i, X_i;\beta_0)] = 0$ である。

$$\begin{aligned} E[s(y_i,X_i;\beta_0)s(y_i,X_i;\beta_0)'|X_i] &= \frac{E\{[y_i - F(X_i'\beta_0)]^2|X_i\}}{F(X_i'\beta_0)^2[1-F(X_i'\beta_0)]^2} f(X_i'\beta_0)^2 X_i X_i' && (66)\\ &= \frac{E(y_i^2|X_i) - F(X_i'\beta_0)^2}{F(X_i'\beta_0)^2[1-F(X_i'\beta_0)]^2} f(X_i'\beta_0)^2 X_i X_i' && (67)\\ &= \frac{1}{F(X_i'\beta_0)[1-F(X_i'\beta_0)]} f(X_i'\beta_0)^2 X_i X_i' && (68)\end{aligned}$$

であり、$E[y_i - F(X_i'\beta_0)|X_i] = 0$ なので

$$E[H(y_i, X_i;\beta_0)|X_i] = -\frac{1}{F(X_i'\beta_0)[1-F(X_i'\beta_0)]} f(X_i'\beta_0)^2 X_i X_i' \tag{69}$$

である。したがって、(iii) が成立する。

(iv) プロビットモデルに関しては面倒である (Newey & McFadden (1994, p.2147, Example 1.2 参照) が、ロジットモデルについては比較的簡単である。$0 \leq \Lambda(X_i'\beta) \leq 1$ と (45) より、すべての β にたいして

$$||H(y_i, X_i; \beta)|| \leq ||X_i X_i'|| \tag{70}$$

となるが、$E||X_i X_i'|| < \infty$ なので (iv) が成り立つ。

(v) $1 - \Lambda(X_i'\beta_0) = \Lambda(-X_i'\beta_0)$ なので、任意の $c > 0$ に対して、

$$\begin{aligned} I(\beta_0) &= -E[H(y_i, X_i; \beta_0)] = E\{\Lambda(X_i'\beta_0)[1 - \Lambda(X_i'\beta_0)]X_i X_i'\} & (71)\\ &\geq E[I(|X_i'\beta_0| \leq c)\Lambda(X_i'\beta_0)\Lambda(-X_i'\beta_0)X_i X_i'] & (72)\\ &\geq \Lambda(-c)^2 E[I(|X_i'\beta_0| \leq c)X_i X_i'] & (73)\end{aligned}$$

最後の行列は c を十分大きくとれば非特異になる。したがって、$I(\beta_0)$ は非特異である。(証明終)

6.5 切断回帰モデルと打ち切り回帰モデル（truncated regression, censored regression model, Tobit model）

もともとのモデルは線形回帰モデルであるが、被説明変数がある閾値以下の個体に関しては観測値が得られない場合（truncated sample）や、ある閾値以下のときは被説明変数が（たとえば非負制約によって）すべて 0 になってしまう場合を考える。例えば、労働供給時間、購入する耐久消費財の価格、といったものが被説明変数である場合に生ずる。その場合、得られたサンプルで OLS 回帰を行うとバイアスが生ずる。なお、説明変数に閾値があって打ち切られる場合はバイアスの問題は生じないため、問題なく OLS 推定ができる。

6.5.1 切断回帰モデル (Truncated regression model)

(y_i, X_i) は次の関係

$$y_i = X_i'\beta + \epsilon_i \tag{74}$$

を満たすが、

$$(y_i, X_i) = \begin{cases} \text{観測可能} & (y_i \geq 0 \text{ のとき}) \\ \text{観測できない} & (y_i < 0 \text{ のとき}) \end{cases} \tag{75}$$

という状況を考える。例えば、女性の労働力市場への参加・不参加の問題を考える際に、個人 i の労働時間を y_i、属性を X_i とすると、上のような構造を考えることができる。また、購入した耐久消費財の価格を説明するモデルにも上のような定式化が考えられる。$y_i < 0$ の個体については観測値がないとき、そのようなサンプルを truncated sample という。このモデルも単純な OLS 推定を行うことができない。ϵ の密度関数、分布関数を f, F とする。

このようなモデルを分析するには、切断のある分布を考える必要がある。u の分布が c より下で切断されているとき、切断された分布は

$$f(u|u > c) = \frac{f(u)}{P(u > c)} = \frac{f(u)}{1 - F(c)} \tag{76}$$

である。$u \sim N(\mu, \sigma^2)$ とすると、切断された分布の期待値と分散は

$$\begin{aligned} E(u|u > c) &= \mu + \sigma\lambda(v) & (77)\\ Var(u|u > c) &= \sigma^2\{1 - \lambda(v)[\lambda(v) - v]\} & (78)\end{aligned}$$

となる。ただし、$v=(c-\mu)/\sigma$、$\lambda(v)=\phi(v)/[1-\Phi(v)]$ で、ϕ,Φ はそれぞれ標準正規分布の密度関数と分布関数である。$\lambda(v)$ は逆ミルズ比 (inverse Mill's ratio)、ハザード関数 (hazard function) と呼ばれる。以下、$y^*|x\sim N(x'\beta,\sigma^2)$ を仮定して OLS と ML 推定法を紹介する。

(1)OLS 推定

$u=y^*$、$c=0$ と考えて上の結果を用いると、切断回帰モデルについて

$$\begin{aligned} E(y_i|X_i) &= X_i'\beta+\sigma\lambda\left(\frac{-X_i'\beta}{\sigma}\right) && (79)\\ Var(y_i|X_i) &= \sigma^2\left\{1-\lambda\left(\frac{-X_i'\beta}{\sigma}\right)\left[\lambda\left(\frac{-X_i'\beta}{\sigma}\right)-\frac{-X_i'\beta}{\sigma}\right]\right\} \end{aligned}$$

を得る。従って、y_i を X_i に回帰する OLS 推定では β を推定できないことがわかる。なぜなら、OLS 推定を行うと、

$$\begin{aligned} E(\hat{\beta}|X) &= \left(\sum X_iX_i'\right)^{-1}\sum X_iE(y_i|X_i) && (80)\\ &= \left(\sum X_iX_i'\right)^{-1}\sum X_i\left[X_i'\beta+\sigma\lambda\left(\frac{-X_i'\beta}{\sigma}\right)\right] && (81)\\ &= \beta+\sigma\left(\sum X_iX_i'\right)^{-1}\sum X_i\lambda\left(\frac{-X_i'\beta}{\sigma}\right) && (82) \end{aligned}$$

となり、バイアスを生ずる。

(2) ML 推定

個体 i の切断前の条件付き密度関数は

$$f(y_i|X_i;\beta,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y_i-X_i'\beta)^2}{2\sigma^2}\right]=\frac{1}{\sigma}\phi\left(\frac{y_i-X_i'\beta}{\sigma}\right) \tag{83}$$

であり、

$$\begin{aligned} P(X_i'\beta+\epsilon_i\geq 0|X_i) &= 1-P(X_i'\beta+\epsilon_i<0|X_i) && (84)\\ &= 1-\Phi\left(\frac{-X_i'\beta}{\sigma}\right) && (85)\\ &= \Phi\left(\frac{X_i'\beta}{\sigma}\right) && (86) \end{aligned}$$

なので、(76) より切断条件付き密度関数は

$$f(y_i|X_i,X_i'\beta+\epsilon_i\geq 0;\beta,\sigma^2)=\frac{\frac{1}{\sigma}\phi(\frac{y_i-X_i'\beta}{\sigma})}{\Phi(\frac{X_i'\beta}{\sigma})} \tag{87}$$

である。従って、観測値 $(y_i,X_i),\ i=1,\cdots,n$ を得たときの対数尤度関数は

$$\log L(\beta,\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-X_i'\beta}{\sigma}\right)^2-\sum_{i=1}^{n}\log\Phi\left(\frac{X_i'\beta}{\sigma}\right) \tag{88}$$

である。これを最大化することにより、MLE が得られる。一般には、パラメータを変換して推定し、また漸近的性質を求めるほうが、簡単である。推定量は一致性と漸近正規性をもつが、その導出は、Hayashi (2000, p515-516) を参照。

6.5.2 打ち切り回帰モデル

$y_i < 0$ の時に、y_i は観測できないが、X_i は観測でき、また $y_i < 0$ と分かっているデータが取得される場合もある。この場合には、打ち切り回帰モデル (censored regression model あるいは、Type 1 Tobit model) という。このモデルに対しても、OLS 推定量はバイアスを生ずる。なお、切断回帰モデルでは、$y_i < 0$ の個体については、(y_i, X_i) が観測されない状況を考えた。

（1） Heckman の 2 段階推定

Heckman (1976) は、回帰に基づく 2 段階推定量を提案した。1 段階目は $y > 0$ の観測値を $y = 1$ で置き換えたデータを用意してプロビット ML で β/σ を推定する。それを $\lambda(-X_i'\beta/\sigma)$ に代入して $\hat{\lambda}_i = \lambda(-X_i'\widehat{\beta/\sigma})$ を作り、2 段階目に y_i が正の観測値のみを用いて (79) に基づいて y_i を $X_i, \hat{\lambda}_i$ に OLS 回帰することによって β の推定量を得る、という方法である。これは CAN 推定量である。詳細は Amemiya (1985, Sec. 10.4.3, p.368-) を参照。

（2） MLE

$y_i^* \geq 0$ の個体に対する対数尤度は

$$-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\left(\frac{y_i - X_i'\beta}{\sigma}\right)^2 \tag{89}$$

で、$y_i^* < 0$ の個体に対する対数尤度は

$$\log\Phi\left(-\frac{X_i'\beta}{\sigma}\right) \tag{90}$$

なので、サンプル全体に対する条件付き対数尤度関数は $n_1 = \sum_{i=1}^n I(y_i > 0)$ として

$$\begin{aligned}
\log L(\beta, \sigma^2) &= -\frac{n_1}{2}\log(2\pi) - \frac{n_1}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^n 1(y_i > 0)\left(\frac{y_i - X_i'\beta}{\sigma}\right)^2 & (91)\\
&\quad + \sum_{i=1}^n 1(y_i = 0)\log\Phi\left(-\frac{X_i'\beta}{\sigma}\right) & (92)\\
&= -\frac{n_1}{2}\log(2\pi) - \frac{n_1}{2}\log\sigma^2 - \frac{1}{2}\sum_{i:y_i>0}\left(\frac{y_i - X_i'\beta}{\sigma}\right)^2 & (93)\\
&\quad + \sum_{i:y_i=0}\log\Phi\left(-\frac{X_i'\beta}{\sigma}\right) & (94)
\end{aligned}$$

これを最大化することによって MLE を得ることができ、一致性をもち、漸近正規で漸近的に有効である。

6.5.3 その他の Tobit model

切断がどのような変数に対して生じるか、何が観測されるかによって色々な Tobit model を考えることができる。

$$\begin{aligned}
y_{1i}^* &= X_{1i}'\beta_1 + \epsilon_{1i} & (95)\\
y_{2i}^* &= X_{2i}'\beta_2 + \epsilon_{2i} & (96)
\end{aligned}$$

で、$*$ のついた変数は観測されない変数とする。

(1) Type 2 Tobit model

$$y_i = \begin{cases} y_{2i}^* & (y_{1i}^* \geq 0 \text{ のとき}) \\ 0 & (y_{1i}^* < 0 \text{ のとき}) \end{cases} \tag{97}$$

(2) Type 3 Tobit model

$$y_{1i} = \begin{cases} y_{1i}^* & (y_{1i}^* \geq 0 \text{ のとき}) \\ 0 & (y_{1i}^* < 0 \text{ のとき}) \end{cases} \tag{98}$$

$$y_{2i} = \begin{cases} y_{2i}^* & (y_{1i}^* \geq 0 \text{ のとき}) \\ 0 & (y_{1i}^* < 0 \text{ のとき}) \end{cases} \tag{99}$$

これらのモデルは原則的に ML 法で推定されるが、他にも実際のデータに合わせて様々なモデル、推定法が提案されている。

References

[1] T. Amemiya. *Advanced Econometrics.* Harvard University Press, Cambridge, Massachusetts, 1985.

[2] F. Hayashi. *Econometrics.* Princeton University Press, 2000.

[3] J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER, 1976.

[4] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1994.

Appendix

定理 1 の証明のための、命題 2 の条件 (iv)' のプロビットの場合の証明

任意の u に対して

$$|\log\Phi(u)| \le |\log\Phi(0)| + |u| + u^2 \tag{100}$$

である。

(証明)

i) $u > 0$ のとき、

$$1 \ge \Phi(u) > \Phi(0) \tag{101}$$

なので、

$$0 \ge \log\Phi(u) > \log\Phi(0) \tag{102}$$

である。したがって、

$$|\log\Phi(u)| < |\log\Phi(0)| \le |\log\Phi(0)| + |u| + u^2 \tag{103}$$

となる。

ii) $u \le 0$ のとき

$$g(u) = \Phi(u) - \Phi(0)\exp(u - u^2) = \Phi(u) - \frac{1}{2}\exp(u - u^2) \tag{104}$$

とおくと

$$\begin{aligned} g'(u) &= \phi(u) - \frac{1}{2}(1-2u)\exp(u-u^2) && (105)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)\left[1 + \frac{\sqrt{2\pi}}{2}(2u-1)\exp\left(u - \frac{u^2}{2}\right)\right] && (106) \end{aligned}$$

である。あらためて、

$$h(u) = 1 + \frac{\sqrt{2\pi}}{2}(2u-1)\exp\left(u - \frac{u^2}{2}\right) \tag{107}$$

とおくと $h(0) = 1 - \sqrt{\pi/2} < 0$、$\lim_{u\to-\infty} h(u) = 1$ である。また、

$$h'(u) = \frac{\sqrt{2\pi}}{2}[2 + (2u-1)(1-u)]\exp\left(u - \frac{u^2}{2}\right) \tag{108}$$

なので、$u \le 0$ の範囲では $h(u)$ は $u = (3-\sqrt{17})/4$ で（負の）最小値を取り、また $u < (3-\sqrt{17})/4$ では単調減少、$u > (3-\sqrt{17})/4$ では単調増加関数である。以上から、ある u^* が存在して、

$$h(u)\begin{cases} < 0 & (u \ge u^*\text{のとき}) \\ \ge 0 & (u < u^*\text{のとき}) \end{cases} \tag{109}$$

が成り立つ。したがって、

$$g'(u)\begin{cases} < 0 & (u \ge u^*\text{のとき}) \\ \ge 0 & (u < u^*\text{のとき}) \end{cases} \tag{110}$$

となる。ここで、$g(0) = 0$ である。

$$\begin{aligned} \lim_{u\to-\infty} g(u) &= \lim_{u\to-\infty}\left[\Phi(u) - \frac{1}{2}\exp(u-u^2)\right] && (111)\\ &= \lim_{u\to-\infty}\left[\frac{\Phi(u)}{\exp(-\frac{u^2}{2})} - \frac{1}{2}\exp\left(u - \frac{u^2}{2}\right)\right]\exp\left(-\frac{u^2}{2}\right) && (112) \end{aligned}$$

で、

$$\lim_{u \to -\infty} \exp\left(-\frac{u^2}{2}\right) = 0, \ \lim_{u \to -\infty} \exp\left(u - \frac{u^2}{2}\right) = 0 \tag{113}$$

$$\lim_{u \to -\infty} \frac{\Phi(u)}{\exp(-\frac{u^2}{2})} = \lim_{u \to -\infty} \frac{\phi(u)}{-u \exp(-\frac{u^2}{2})} = 0 \tag{114}$$

なので、$\lim_{u \to -\infty} g(u) = 0$ である。したがって、$u \leq 0$ のとき $g(u) \geq 0$ である。

したがって、対数関数の単調性より

$$\log \Phi(u) \geq \log \Phi(0) + u - u^2 \tag{115}$$

$$-\log \Phi(u) \leq -\log \Phi(0) - u + u^2 \tag{116}$$

また $\log \Phi(u) \leq 0$, $\log \Phi(0) \leq 0$, $u \leq 0$ なので

$$|\log \Phi(u)| \leq |\log \Phi(0)| + |u| + u^2 \tag{117}$$

となる。(証明終)