第5講　質的従属変数パネルデータ分析

5.1　はじめに

ミクロ経済学の多くは選択問題を扱っている。消費者の消費財の選択、学校や就労先の企業の選択、居住地・住宅の選択、結婚・離婚の選択、出産の決定、退職の決定、再就職先の選択、保険への加入、証券投資の決定、企業の市場への参加の決定、企業の投資決定、企業の雇用決定、年金や退職給付制度の決定、公共財の供給決定、国際経済学の問題に入る通貨圏への参加の決定、貿易協定の締結など数えればきりがない。

近年、ミクロ経済学がゲーム理論で説明されるようになってきたが、これが可能になったのはミクロ経済学の多くの問題がゲームのように次の動きを選択するという形で設定されているからであると言える。これらの選択問題あるいは意志決定問題を実証的に分析してみるということはミクロ経済理論を検証するという意味で極めて重要なことであり、かつ現実的にも興味深いものである。

実証研究において、選択問題を扱おうとすると、選択した場合を 1、選択しなかった場合を 0 とおくことで、本質的には質的な情報を数量化して，選択行動を統計的にモデル化することが可能となる。

このアプローチの拡張として、選択すれば任意の正の数値をとるが、選択しなければ 0 であるという場合が考えられる。例えば、株式への投資は、投資を選択しなければ 0 にとどまるが、選択すればあとは投資家の投資額は個々人で違ってくる。選択した後の量の決定が任意の場合には、このようなアプローチが有益になってくるのである。

これまでこのような選択問題は主としてクロスセクション・データを用いて実証されてきたし、その統計手法もクロスセクション・データを中心に開発されてきた。当然予想されるようにパネルデータを用いれば、経済主体の意思決定の問題はより精緻に分析することができる。しかし、同時に、クロスセクション・データの選択モデルがすでに非線形モデルであり、その推計はかなり複雑になっているのだが、それをパネル・データに拡張するためには、かなり強い制約をおく必要がでてくる。

現在のところ、クロスセクション・データを用いた分析方法がある多項反応データを用いた順序プロビット、多項ロジット、ネステッド・ロジットなどのパネルデータ分析への応用はまだ時間がかかると思われる。これらのパネルデータ分析への拡張は将来の課題としておきたい。また、パテントの数や事故数などを数え上げるタイプのデータであるカウントデータのパネル分析については手法は開発されているが、本書では扱わない[1]。

[1]カウントデータのパネルデータ分析については Cameron and Trivedi (1998) や Winkel-

5.2 質的従属変数クロスセクション推定

これまでクロスセクションデータでよく用いられてきた質的（離散的）従属変数を用いた推定はパネルデータでも有効である[2]。具体的に例を挙げれば、車を買うかどうか、あるいは車を所有しているかどうか、住宅を買うかどうか、労働組合に参加するかどうか、結婚するかどうかなどの意思決定問題に計量経済学的に答えることができる。既に述べたように、従属変数 y_{it} は一般に選択しなければ 0、選択すれば 1 の 2 項選択の形をとることが多いが、経済主体 i が時間 t に意思決定をする（例えば、結婚する）確率を p_{it} と表せば、従属変数の期待値は $E(y_{it}) = 1 \cdot p_{it} + 0 \cdot (1 - p_{it}) = p_{it}$ となり、これは通常、なんらかの変数 (x_{it}) で説明される[3]。

$$p_{it} = \Pr[y_{it} = 1] = E(y_{it}|x_{it}) = F(x_{it}) \tag{1}$$

クロスセクションデータを用いた実証研究では $F(x_{it})$ の定式化としてプロビット・モデルとロジット・モデルがそれぞれ次のように定義されている。

両端で 0 と 1 をとる単調な関数として逆関数 $\Phi^{-1}(x)$ に関する線形モデルを考える。

$$\Phi^{-1}(x_{it}) = \frac{x_{it} - \bar{x}}{\sigma} \tag{2}$$

この累積密度関数は

$$F(x_{it}) = \Phi(x_{it}) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \tag{3}$$

と表せる。これは標準正規分布の密度関数に他ならず、このようなモデルを**プロビット・モデル**と呼ぶ。

また同様にしてロジスティック分布の密度関数を考えることもできる。

$$F(x_{it}) = \frac{e^{\mu_0 + \beta_1 x}}{1 + e^{\mu_0 + \beta_1 x}} \tag{4}$$

これを**ロジット・モデル**と呼ぶ。

これらのモデルでは、実際に何らかの意思決定がなされたとすると、従属変数が直接は観察できないある水準を超えたことを意味している。すなわち、

$$\begin{aligned} y_{it} &= 1 \quad if\ y^*_{it} > 0 \\ y_{it} &= 0 \quad if\ y^*_{it} \leqslant 0 \end{aligned} \tag{5}$$

ここで、$y^*_{it} = \mu_0 + \beta_1 x_{it} + v_{it}$. つまり

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[u_{it} > -\mu_0 + \beta_1 x_{it}] = F(\mu_0 + \beta_1 x_{it}) \tag{6}$$

となる。

---

mann (2003) などで扱われているが、著者の知る限り実証での応用はまだ少ない。

[2] この分野における基本文献は Maddala(1983, 1987) である。また、最近の文献には Gourieroux (2000)、Lee (2002) がある。以下では東京大学教養学部統計学教室（編）（1992、第 8 章）を参照している。

[3] クロスセクション・データでは t が固定されており、パネル・データでは t が変動していると考えればよい。

ところで

$$P_{0i} = \Pr(y_{it} = 0 \mid x_{it}), \qquad P_{1i} = \Pr(y_{it} = 1 \mid x_{it})$$

とすると、

$$P_{0i} = 1 - P_{1i} = \frac{1}{1 + \exp(\mu_0 + \beta_1 x_{it})}$$

であるから、

$$\log(P_{1i}/P_{0i}) = \mu_0 + \beta_1 x_{it}$$

となる。ロジット・モデルは y_{it} が 0 か 1 か 2 つの確率比 (オッズ) の対数 (対数オッズ) が x_{it} の線形式であると想定していることになる。

標本 $\{y_{it}\}$ を得る同時確率は次のような尤度関数として表せる。

$$L(\mu_0, \beta_1) = \prod_{y_i=1} F(\mu_0 + \beta_1 x_{it}) \cdot \prod_{y_i=0} (1 - F(\mu_0 + \beta_1 x_{it})) \tag{7}$$

上式両辺の対数をとると次のような線形式に変換できる。

$$\log(\mu_0, \beta_1) = \sum_{i=1}^{n} \{y_{it} \log F(\mu_0 + \beta_1 x_{it}) + (1 - y_{it}) \log(1 - F(\mu_0 + \beta_1 x_{it})\} \tag{8}$$

これを最大化する推定量は一階条件より求められる。

$$\frac{\partial \log L(\mu_0, \beta_1)}{\partial \mu_0} = 0, \quad \frac{\partial \log L(\mu_0, \beta_1)}{\partial \beta_1} = 0 \tag{9}$$

この解は関数 F の形状に依存するが、いずれにしても $L(\mu_0, \beta_1)$ は単調な凹型関数であり、最適解は一般に存在する[4]。最尤推定量 $\mu_0, \hat{\beta}_1$ は一致推定量であり、標本の大きさが増加するに従って μ_0, β_1 の真の値に近づく。不偏性は一般には成り立たない。

5.3 パネル・プロビット・モデルとパネル・ロジット・モデル

5.3.1 モデルの構造

パネルデータの場合、誤差項に固定効果 μ_i が入ることで従来のプロビット分析、ロジット分析とは異なってくる。次のような固定効果モデルを考えよう。

$$y^*_{it} = x^{'}_{it}\beta_1 + \mu_i + \nu_{it} \tag{10}$$

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[\nu_{it} > -x^{'}_{it}\beta_1 - \mu_i] = F(x^{'}_{it}\beta_1 + \mu_i) \tag{11}$$

[4] 最適解が β_1 無限大となる時には最尤推定量は存在しない。

ここで、T を固定すると、固定効果 μ_i のパラメータは N に応じて増加する。これは、パラメータ μ_i は固定された T に対して一致推定を得ることが出来ないことを意味している[5]。線形パネルデータ回帰モデルでは、パラメータ μ_i はウィズイン推定によって除去して、β だけに対しては一致推定を得ることができる[6]。

この問題の直感的な説明は次のようにできる。クロスセクションデータの場合、行動に変化を与える閾値 y^* は標本全体で共通の値が想定されており、$y_{it} = \mu_0 + \beta_1 x_{it} + v_{it}$ と設定した場合のパラメータの関係は図表 5.1 のようになる。

図表 5.1　ロジスティック分布関数（μ_0 共通の場合）

閾値 y^* が所与の場合、$\mu_0 = 0$ で共通となり、関数のフィットは β_1 の値によって決まってくる。これは最尤法によって β_1 を決定することによって一致推定を得ることができる。

ところがパネルデータの場合、閾値 y^* は個別主体 i 毎に異なり得る。図表 5.2 で表したようにパラメータ (μ_i, β_1) において、β_1 を共通であるという制約下で N が増加するにつれて μ_i の一致推定を得ることは難しくなる。また与えられた T 期間内に各経済主体が 0 から 1 へ確率を変化させるとは限らない。期間中一貫して 0 あるいは 1 をとりつづける主体もあるのでその場合には閾値 y^* や固定効果 μ_i を確定することはできない。

クロスセクション推計であれば $\mu_i = \mu_0$ で全てのデータに対し共通であったが、パネル推計では個別 μ_i が成り立ち、しかも同時に推計するパラメータ β_1 は全てのデータに共通であると想定されている。この場合、μ_i を付随（incidental）パラメータと呼び β_1 を構造（structural）パラメータと呼ぶ。

問題は、このような μ_i が確定 (識別) できるかということと、その場合 β_1 の推定にバイアスはないかということである。

図表 5.2　ロジスティック分布関数（μ_i が可変の場合）

5.3.2　付随パラメータ問題

一般的な線形パネルモデルを考える。

$$y_{it} = \mu_i + \beta x_{it} + v_{it}$$

[5] この問題は Neyman and Scott (1948) によって古典的付随パラメータ問題 (the classical incidental parameter problem) と呼ばれているものに相当する。Lancaster (2000) はこの問題は現在も解決されていないことを指摘した上で、固定効果の直交条件を見つけることが重要であると指摘している。

[6] Hsiao(2003) でも示されているように、β と μ_i が漸近的に独立であれば、線形モデルの最尤法で β の一致推計を得ることが出来る。これが非線形モデルの場合やプロビット・モデルの場合には一致推計を得ることが出来ない。

付随パラメータ μ_i と構造パラメータ β がともに適切に推計できれば問題はないが、誤差構成から μ_i が直交条件によって分離できなければ β の推計は不一致となる。

条件付け分離による解決

もし μ_i と β を条件付け関数として 2 つの尤度関数に分離できるとする[7]。

$$L(y \mid \mu, \beta) = L_1(s \mid \mu) L_2(y \mid s, \beta) \tag{12}$$

尤度関数 L_1 が L_2 から完全に独立（likelihood orthogonal）であれば、μ_i, β は一致推計となる。すなわち

$$\frac{\partial^2 \log L(\mu, \beta)}{\partial \mu \partial \beta} = 0 \tag{13}$$

L_1 と L_2 が完全には独立していない場合

$$L(y \mid \mu, \beta) = L_1(s \mid \mu, \beta) L_2(y \mid s, \beta) \tag{14}$$

s は μ には依存していないが、β には依存している。この場合でも、尤度関数 L_2 を最大化するような β を得ることは可能である。

尤度関数が次のように分離できる場合

$$L(y \mid \mu, \beta) = L_1(s \mid \beta) L_2(y \mid s, \mu, \beta) \tag{15}$$

s の限界分布に基づいて、β と μ を推計することが可能である（これを partial likelihood procedure と呼ぶこともある)。

もし μ_i が固定効果であり、β が共通パラメータであるとき、$\mu_i = \mu(\mu_i, \beta)$ と表せるとする。[8]ここで μ_i は β と直交しており、μ_i の代理パラメータと考える。

次のような関係が導ければ

$$\frac{\partial^2 \log L_i}{\partial \mu_i \partial \beta} = 0 \tag{16}$$

ここで L_i は個別主体 i の尤度貢献分であると考えられる。対数尤度関数を μ^* と β について微分して、期待値をとると

$$\frac{\partial \mu_i}{\partial \beta_j} E\left(\frac{\partial^2 \log L_i}{\partial \mu_i^2}\right) + E\left(\frac{\partial^2 \log L_i}{\partial \mu_i \partial \beta_j}\right) = 0 \quad j = 1, \ldots, k \tag{17}$$

を得る。これを解くと μ_i が β に依存していることが明らかになる。直交条件付固定効果 μ_i を用いて β を最尤法推定するというのが一つの解決策である。具体的には次のような尤度関数を考える。

$$L_M(\beta) = L(\beta, \mu^*_\beta) |j_{\mu^* \mu^*}(\beta, \mu^*_\beta)|^{-\frac{1}{2}} \tag{18}$$

[7]Lancaster(200; pp.400-402) 参照。ここで s はある統計量で、α と β を分離する (cut) 際の指標となる。例えば、s は各個人が 1 の値をとった回数であるとか、各時点において 0 の値をとったクロスセクションの合計数とかを用いる。

[8]Cox and Reid (1987) を参照。

ここで $j_{\mu^*\mu^*}(\beta,\mu^*) = -\frac{\partial^2 \log L(\beta,\mu^*)}{\partial\mu^*\partial\mu^*}, j_{\mu^*\mu^*}$ は個別対数尤度関数の Hessian 行列に負の符号をつけたのものである。また $L(\beta,\mu^*_\beta)$ は β の最尤推定を与えるような尤度関数である。

付随パラメータ問題は未解決であるが、Chamberlain (1980) は μ_i の最小十分統計量は $\sum_{t=1}^T y_{it}$ であることを示し、次のような条件付尤度関数を最大化して β のロジット推定を得ることを提唱した。

$$L_c = \prod_{i=1}^N \Pr(y_{i1},....y_{iT} \mid \textstyle\sum_{t=1}^T y_{it}) \tag{19}$$

この方法では十分統計の定義により、推定されたパラメータ β は μ_i に依存しない。

パネルロジットの場合 P_r は次のように表せる。

$$P_r(y_{it}=1 \mid \mu_i,\beta,x_{i1},\ldots,x_{iT}) = \frac{1}{1+e^{-\mu_i-\beta' x_{it}}} \tag{20}$$

μ は β と直交していない。β と直交しているような代理になる固定効果を次のように定義する。

$$\mu_i = \sum P_r(y_{it}=1 \mid \mu_i,\beta,x_{i1},\ldots,x_{iT}) \tag{21}$$

この最尤推定量は

$$\mu_i = \textstyle\sum_{t=1}^T y_{it} \tag{22}$$

となる。この μ_i を用いて推計される β は一致推定量となる。ベイジアン推計として μ_i に相当する prior(事前情報) が利用可能かどうかは不明である。

このモデルは Chamberlin の提示した条件付最尤法と固定効果を考慮しない通常のロジット推定の差を Hausman の χ^2 検定の要領で検定できる。通常のロジット推定が有効一致推定であるのは固定効果がない場合であり、固定効果がある場合には一致推定にはならない。Chamberlin の推定は固定効果の有無にかかわらず一致しているが、固定効果がない場合には、有効ではなくなる[9]。

代替的なモデルとして Liang and Zeger (1986) が提案した Generalized Estimating Equations (GEE) Population-averaged Model がある。これは、ランダム効果線形推定法であり、プロビット推定法を線形近似した簡便法である[10]。

[9] これに対して、固定効果プロビット・モデルでは計算はロジット・モデルのように簡単ではない。一般に固定効果を含んだ最尤法は、N が大きく、T が固定されている場合には、一致推計量が得られない。Heckman (1981b) を参照。Manski(1987) は条件付最尤法が適用できない場合に固定効果の下で一致推定パラメータ β を最大スコア推定法（Manski(1975,1985) 参照）によって得られることを示した。しかしこの推定方法は収束に時間がかかり、漸近的に正規化しないと言われている（Kim and Pollard (1990)）。

[10] STATA コードでは xtgee というコマンドで推定できる。

5.4 クロスセクション・トービット・モデル

トービット･モデルは Tobin (1958) で初めて論じられ、Goldberger (1964) によってトービットと命名され、一般に広く紹介されるようになった。Amemiya(1985, Chapter 10) や Maddala(1983,Capter 6) で包括的にサーベイされている。

このモデルは、例えば、所得が一定水準以上に達すると、海外旅行をすると考えられる時に、一定水準以下の所得の家計も含めた全サンプルの海外旅行の需要関数の推定に用いることができる。この需要の確率分布のイメージは図表 5.3 のように表せる。

図表 5.3　C で切断されたデータの分布

このように分布の端が切断されたようなデータを用いる場合には、通常の最小二乗法で推計すると、誤差項が正規分布せずに推計パラメータにバイアスをもたらす。この問題を解決するために提案されたのがトービット・モデルである。

トービット・モデルは次のように定式化されている[11]。

$$y_i^* = x_i^{'}\beta + u_i \quad u_i \sim iidN(0, \sigma_u^2) \tag{23}$$

$$\begin{aligned} y_{it} &= y_{it}^* \quad if \; y_{it}^* > 0 \\ y_{it} &= 0 \quad if \; y_{it}^* \leq 0 \end{aligned} \tag{24}$$

ここで $y_i > 0$ に含まれる人数を N_1、$y_i = 0$ に含まれる人数を N_0 とする。ここでまず、N_1 が決定されるとして、その場合の y_i の密度関数と分布関数をそれぞれ次のように定義する。

$$F_i = F(\beta' x_i, \sigma^2) = \int_{-\infty}^{\beta' x_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-t^2/2\sigma^2} dt \tag{25}$$

$$f_i = f(\beta' x_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2\sigma^2)(\beta' x_i)^2} \tag{26}$$

$y_i = 0$ となる確率は次のように定義できる。

$$\Pr ob(y_i = 0) = \Pr ob(u_i < -\beta' x_i) = (1 - F_i)$$

$y_i > 0$ となる確率は次のように定義される。

$$\Pr ob(y_i > 0) \cdot f(y_i | y_i > 0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2\sigma^2)(y_i - x_i'\beta)^2}$$

この二つの確率を掛け合わせて同時確率を求める尤度関数は次のように書ける。

[11] 以下では Amemiya(1985,Chapter 10) および Maddala(1983,Chapter 6) を参照。

$$L = \prod_{N_0}(1-F_i)\prod_{N_1}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(1/2\sigma^2)(y_i - x_i'\beta)^2} \tag{27}$$

尤度関数を最大化することによってパラメータ β を求めることができる。

$$\beta = \beta_{OLS} - \sigma(X_1'X_1)^{-1}X_0'\gamma_0 \tag{28}$$

ここで β_{OLS} はサンプル N_1 に含まれる $y(>0)$ を最小二乗推定して得たものである。これは最尤法推定と最小二乗推定の関係を示している。

図表 5.4　トービット・モデルの類型

Amemiya (1985) はトービット・モデルをさらに５つのタイプに分類し、それぞれの尤度関数を導き、また当該タイプを用いた実証研究をサーベイしている。これは、トービット・タイプのモデルを分析する上で極めて有益な分類であり、多くの研究者が Amemiya（1985) の分類に従っている[12]。

簡単に、それぞれのタイプの特徴を要約しておこう。

（1）タイプ I は標準トービット・モデルと呼ばれるもので、既に説明した通りである。データのタイプによって　途中打ち切り回帰モデルと　切断回帰モデルに分けることができる。途中打ち切りデータとは従属変数が負の値をとれないような時に得られるデータであり、切断データとはある値を超えるようなサンプルは調査しないような時[13]に得られるデータである。

（2）タイプ II は２つの観察不可能な従属変数を考慮している。図表 5.4 の記号に従えば、y_1 は正負符号だけが観察され、y_2 は y_1 が負の場合は観察されない変数である。このモデルは例えば、Gronau (1973) や Mroz (1987) など労働経済学では広範に用いられてきた。例えば、Gronau (1973) では女性の時間の機会費用と市場賃金を考えて、女性が労働市場に参入した場合のみ市場賃金が観察できるという状況に応用している。このモデルを最尤法で推定することは可能であるが、必ずしも収束しなかったり、最尤値に到達しているかどうか疑わし場合があり、現実的には問題がある。そこでタイプ III で紹介するようにヘックマンの２段階推定法が簡便な代替的推定法として広く用いられている。

（3）タイプ III では、y_1 は正の場合観察され、y_2 も y_1 が正の場合観察され、負の場合は観察されない変数である。タイプ II と比べれば、情報量は多い。このタイプは Heckman (1974) で論じられている女性の労働供給・賃金関数モデルが該当する。ヘックマンは Gronau のモデルに労働時間を加えて、賃金と労働時間が同時決定されるモデルを考察し、プロビット法と最小二乗法によって推定できる２段階推定法を提示した。

$$y_i = \mathbf{x}_i'\beta + \sigma\lambda(\mathbf{x}_i'\alpha) + \varepsilon_i \quad y_i > 0 \tag{29}$$

[12]例えば、Honoré, Kyriazidou and Udry (1997) や縄田（1997, 2003) などを参照。

[13]例えば就労している女性だけを対象にした調査であるとか、所得が１億円以上の家計は調査対象としない場合など、さまざまな場合が考えられる。

ここで $\alpha = \beta/\sigma$、$E(\varepsilon_i) = 0$、$\lambda(\mathbf{x}) = f(\mathbf{x})/F(\mathbf{x})$ でミルズ比 (Mill's ratio) の逆数である。

このとき誤差項の分散は次のように表せる。

$$Var(\varepsilon_i) = \sigma^2 - \sigma^2 \mathbf{x}_i' \alpha \lambda(x_i' \alpha) - \sigma^2 \lambda(\mathbf{x}_i' \alpha)^2 \tag{30}$$

ヘックマンは次の２段階の推定法を提案した。

第１ステップ：パラメータ $\hat{\alpha}$ をプロビット最尤法で推定する。

第２ステップ：パラメータ $\hat{\alpha}$ を (29) 式に代入し、 $y_i > 0$ のサンプルについて最小二乗推定し、パラメータ $\hat{\beta}$ を得る。これは一致推定量である。

この場合、White の robust 推定を行って不均一分散の問題に対処しておくべきである。ヘックマンの２段階推定法[14]は広く使われているが、縄田（1997）が指摘しているように、第一段階のセレクションで用いる説明変数 x_{1i} と第２段階の推定で用いる説明変数 x_{2i} がかなり重複しているような場合には、多重共線性の問題がおこりヘックマンの方法は最尤法に比べて、誤差の分散が大きくなり望ましくない。逆に x_{1i} と x_{2i} が別の変数である場合には、最尤法と比べても誤差分散は大きくなく、推定方法の簡便さを考えると、ヘックマンの方法を用いることが正当化される。

（4）タイプ IV はタイプ III に $y_1 > 0$ の場合、変数 y_2 が観察されるのに加えて、$y_1 \leqq 0$ の場合に観察される変数 y_3 を考慮したモデルである。

推定方法はタイプ III とほぼ同様で、ヘックマンの２段階推定法を用いることができる。このタイプの応用例は Kenny, Lee, Maddala and Trost (1979) による大学教育の所得への効果を最尤法で推定している。Nelson and Olson (1978) も教育の所得への効果を推定している。Tomes (1981) は遺産と相続人の所得を同時方程式で推定している。

（5）タイプ V はタイプ IV で y_1 の正負符号だけが観察される場合を扱っている。推定方法はタイプ IV より幾分簡単になっており、特にこのタイプ特有の推定上の問題はない。このタイプは Lee(1978,1981) が考察している。Lee(1978) では組合に入った場合の賃金（y_2）と入らなかった場合の賃金（y_3）にどちらが得になるかを正負符号だけ観察可能な y_1 とおいて、Heckamn の２段階推定法を用いて推計している。しかし、y_1 は同時決定されるはずであり、同時性の問題はここでは無視されている。Heckamn(1978b) では同時方程式のタイプ V モデルが論じられている。

5.5 パネル・トービット・モデル

パネルデータでは固定効果トービット・モデルは次のように定義できる。

$$y_{it}^* = x_{it}^{'}\beta + u_{it} \quad u_{it} = \mu_i + \nu_{it} \quad \nu_{it} \sim iidN(0, \sigma_\nu^2) \tag{31}$$

[14]STATA では heckman というコマンドを用いることでヘックマンの２段階推定法を行うことができる。また heckprob というコマンドでセレクション用のプロビット最尤法推定ができる。

$$y_{it} = y_{it}^* \quad if\ y_{it}^* > 0 \tag{32}$$
$$y_{it} = 0 \quad if\ y_{it}^* \leq 0$$

ここで $d_{it} = 1\ \ if\ y_{it}^* > 0,\ d_{it} = 0\ \ if\ y_{it}^* \leq 0$ とすると、対数尤度関数は次のように定義できる。

$$\log L = \sum_{i,t}(1-d_{it})\log\Phi(\frac{-x_{it}\beta-\mu_i}{\sigma})+\sum_{i,t}d_{it}\{-\frac{1}{2}\log\sigma^2-\frac{1}{2\sigma^2}(y_{it}-x_{it}^{'}\beta-\mu_i)^2\} \tag{33}$$

線形モデルとは違い β と σ は μ_i に依存する。これまで何度も論じてきたように、パラメータ μ_i は固定された T に対して一致推定を得ることが出来ない。この不一致はパラメータ β と σ を通して発生する。

Heckman and MaCurdy (1980) は反復法 (iterative methods) によって推定することを提唱した[15]。すなわち、β と σ に対して初期値を与え、それを所与として、上述の対数尤度関数を μ_i に関して最大化する。その値を再び尤度関数に代入し、今度は β と σ に関して最大化し、新たな β と σ を得る。この作業を β と σ が収束するまで繰り返すのである。

Honoré(1992) は誤差項の分布を特定化しないセミパラメトリック推定を提唱している[16]。具体的には（23）式より次のように定義する。

$$u_{ist}(b) = \max\{y_{is}, (x_{is}-x_{it})b\} - \max\{0, (x_{is}-x_{it})b\} \tag{34}$$

$b = \beta$ の場合、

$$\begin{aligned} u_{ist}(\beta) &= \max\{y_{is}, (x_{is}-x_{it})\beta\} - \max\{0, (x_{is}-x_{it})\beta\} \\ &= \max\{\mu_i + \nu_{it}, -x_{is}\beta, -x_{it}\beta\} - \max\{-x_{is}\beta, -x_{it}\beta\} \end{aligned} \tag{35}$$

ここで $u_{ist}(\beta)$ は s と t に関して対称である。ν_{it} が $i.i.d.$ に従っているとすれば、$u_{ist}(\beta)$ と $u_{its}(\beta)$ も $i.i.d.$ に従う。このことから、次のモーメント条件が導かれる。

$$E[(\xi(\psi(u_{its}(\beta)) - \psi(u_{ist}(\beta))))|x_{it}, \mu_i] = 0 \tag{36}$$

この式は次のように書き換えることが出来る。

$$E[(\xi(\psi(u_{its}(\beta)) - \psi(u_{ist}(\beta))))(x_{it}-x_{is})] = 0 \tag{37}$$

これはさらに次のように変換できる。

$$E[r(y_{is}, y_{it}, (x_{it}-x_{is})\beta)(x_{is}-x_{it})] = 0 \tag{38}$$

[15]彼らは、従属変数のラグが説明変数に入っていないのならば、μ_i の一致推計を得ることが出来ないということはそれほど大きな問題ではないと述べている（Heckman and MaCurdy(1980, P.59))。

[16]以下は Honoré (1992) の他、Honoré and Kyriazidou (2000)、Arellano and Honoré (2001, pp.3272-75)、Hsiao (2003, pp.243-59) を参照している。

ここで $r(.,.,.)$ は第 3 変数に対して単調関数であり、第 3 変数に関して r を積分して、モーメント条件を最小化問題に変換する。

$$\min_b E[R(y_{is}, y_{it}, (x_{is} - x_{it})b)] \tag{39}$$

パラメータ β は上式を最小化することで s と t のペアに対して刈込最小二乗推定 (trimmed least squares estimator) ができる。これは一致推定であり、$\sqrt{n}$ の速度で漸近的に正規化する。$\xi(d) = sign(d)$、$\psi(d) = d$ と定義すると、パラメータ β は次のように表せる。

$$\hat{\beta} = \arg\min_b \textstyle\sum_i^N (1 - 1\{y_{i1} \leqslant \triangle x_i b, y_{i2} \leqslant 0\}) \tag{40}$$
$$\times(1 - 1\{y_{i2} \leqslant - \triangle x_i b, y_{i1} \leqslant 0\})|y_{i1} - y_{i2} - \triangle x_i b|$$

このアプローチの意義は次のように要約できる。個人のデータ切断点が異時点で違う場合には、階差をとっても固定効果を消去することができないということから、２時点のペアを作り、切断点の大きい方にそろえて、データを刈り込めば、残ったデータの分布の形状は一致するので一致パラメータ β を推定することができるということである。問題はこの方法ではデータを大幅に捨ててしまうことがあり、また、切断点が時間とともに変化すること自体に意義があり、それを強引に揃えることにも問題がある。いずれにせよ、このアプローチはまだ実証で広く使われるには至っておらず、今後の評価を待っているところである[17]。

5.6 ダイナミック・トービット・モデル

前節で解説したパネル・トービットでさえ、かなり強引な制約をおかない限り一致推定が行えなかったが、被説明変数のラグが説明変数に加わったダイナミック・トービット・モデルではさらに状況が複雑である。とりわけ厄介な問題はラグとして入ってくる変数の初期値と固定効果の関係である。データ自体が実際に経済行動の初期から観察されていれば、初期値問題は無視できるが、すでに歴史が進展している途中でたまたま観察されたデータを初期値として扱うと、固定効果と複雑に関係しており一致有効推定を得ることは難しい。

ダイナミック・トービット・モデルは次のような構造をしている。

$$y^*_{it} = \gamma y^*_{it-1} + x'_{it}\beta + u_{it} \quad u_{it} = \mu_i + \nu_{it} \quad \nu_{it} \sim iidN(0, \sigma^2_\nu) \tag{41}$$

[17] Honoré (1992) で用いられた推定方法の GAUSS プログラムは Honoré のホームページで公開されている（http://www.princeton.edu/~honore/pantob)。関心のある方は利用されたい。また、Kyriazidou (1997) は、固定効果があるタイプ２トービット・モデルをサンプル・セレクションが内生的に決まるようなモデルに拡張して、その推定方法を検討している。その GAUSS プログラムは彼女のホームページで公開されている（http://www.econ.ucla.edu/kyria)。

$$y_{it} = y_{it}^* \quad if\ y_{it}^* > 0 \tag{42}$$
$$y_{it} = 0 \quad if\ y_{it}^* \leqq 0$$

ここでは被説明変数のラグは１であるが、複数ラグに拡張できる。

Honoré(1993) は前節と同様にペアの刈込みを行いデータの対称性を回復させた上で、モーメント条件を導いた。

$$E[\psi(u_{ist}(\beta,\gamma)) - \psi(u_{its}(\beta,\gamma))|\{x_{it}\}_{t=1}^T] = 0 \tag{43}$$

上式を満たすようなパラメータ (β,γ) を GMM 推定することを提案している。Honoré and Hu (2004) は (43) 式が成り立つための十分条件を導いている。また Hu(1999) によれば、線形 GMM 推定と上述の切断データに対応した非線形 GMM 推定とでは推定結果に有意な違いが見られることから、この推定方法の有効性を論じている。

Arellano, Bover and Labeaga (1999) は一部観察不可能なデータを含んだダイナミック・トービット・モデルを線形自己回帰２段階推定で解く方法を提案している。この方法は、第一ステップで y_{it}^* を過去全てのラグ項、y_{i0}^*、y_{i1}^*、$y_{i2}^* \cdots .y_{i\ \mathrm{t}-1}^*$ と x_{it}、x_{it}、x_{i3}、$\cdots x_{it}$ で説明するモデルから推定する、第２ステップで（41) 式を最小距離法（Minimum Distance Estimator）で非線形推定する、というものである。第１ステップで全ての過去について $y_{it-s} > 0$ を満たす個人が選択されたという仮定は強すぎる上に、第２ステップの非線形推定も簡単には導出できないなどの問題がある。

Kyriazidou (2001) は Kyriazidou (1997) をダイナミックな選択モデルに拡張し２段階 GMM 推定（kernel-weighted GMM:KGMM）を提案している。

前節末で述べたように、これらの様々な推定方法は理論的正当性と同時に計算の簡便性、収束スピードなど実証上のパフォーマンスに基づいて評価されなければならない。現状では、ここで紹介した推定方法はまだ実証研究のツールとして確立されたものにはなっていない。

5.7 STATA コード

質的従属変数パネルデータ分析は現在最も活発に研究が進んでいる分野であり、新しい推定方法が次々と提案されているが、STATA で利用可能なコマンドはパネル・ロジット推定（xtlogit)、パネル・プロビット推定 (xtprobit)、パネル・トービット推定 (xttobit) の３種類である。もちろん研究者が個別に書いたプログラムで公開されているものも多いし、STATA コードで書かれたものもある。関心のある方はそれらのプログラムを使ったり、自ら STATA コードに転換したりすることをお勧めする。

ここでは Wooldridge (2003) が教育目的で公開しているデータ (http://www.msu.edu/˜ec/wooldridge/book2.htm) に入っている KEANE.DTA というデータを使ってパネル・ロジット推定、パネル・プロビット推定を行う。このデータはもともと、Kean and Wolpin (1997) で用いられたデータの一部

を教育用に Wooldridge が用いているものである。Kean and Wolpin (1997) は既に紹介した National Longitudinal Survey of Labor Market Experience (NLSY) の 1979 年コーホートのデータを用いて教育、就業選択に関するパネルデータ分析を行っている。ここでは Wooldridge (2002, pp.498-499) が 1987 年のクロスセクション・データについて検討したモデルを 1981-87 年のパネル・データに拡張してパネル・ロジット分析、パネル・プロビット分析を行った。推計したモデルは次のようなものである。

$$enroll = \alpha_i + \beta edu_{it} + \gamma(\exp er)_{it} + \delta(\exp er)^2_{it} + \eta black_i + \varepsilon_{it}$$

ここで $enroll$ は学校へ行って教育をうける場合に 1 となり、それ以外で 0 となる変数、edu は過去の教育年数、$exper$ は労働経験年数を表わしている。$expersq$ は労働経験年数の 2 乗、$black$ は黒人であれば 1、それ以外は 0 となる変数である。ここで問題にしているのは社会人が追加的に教育を受けるインセンティブはどれぐらいあるかということである。元のデータには無職在宅 (home) や就労（work）などの変数もあり、引きこもりや就労インセンティブなど様々な問題が分析できるが、ここでは追加的教育を選択した人のモデルを考えている。STATA コードは次の通りである。

```
/*Panel Logit*/
xtlogit enroll educ exper expersq, fe
xtlogit enroll educ exper expersq black, re

/*Panel Probit*/
xtprobit enroll educ exper expersq black, re
```

結果は図表 5.5 に載せてある。容易に想像がつくように、高い教育を受けた人、労働経験の長い人ほど追加的な教育は受けたがらないことが分かる。パラメータの値は多少違うが、傾向はロジットもプロビットも同じである。尤度比検定の結果はプーリング・データではなくパネル・データを使った方がいいことを示唆している。

次いで、パネル・トービット分析についても Wooldridge (2002) が教育目的で公開しているデータ (http://www.msu.edu/~ec/wooldridge /book2.htm) に入っている JTRAIN1.DTA というデータを使う。このデータはミシガン州の製造業企業の労働者の生産性向上のための職業訓練に関する 1987-89 年のパネルデータである。ここではこのような職業訓練を施している企業の傾向を見つけようというのが目的である。推計モデルは次のようなものである。

$$\ln hrsemp_{it} = \alpha_i + \beta \ln employ_{it} + \gamma union_{it} + \delta grant_{it} + \varepsilon_{it}$$

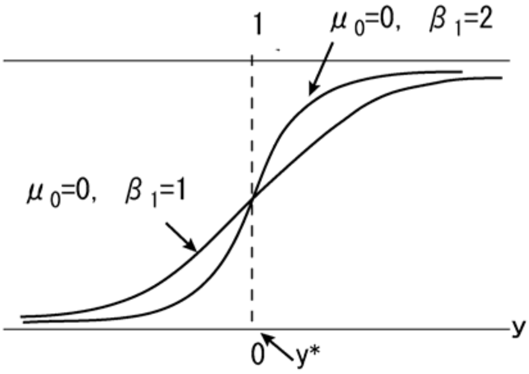
ここで $\ln hrsemp = \log(1 + hrsemp)$、$hrsemp$ は一人当たりの訓練時間、$employ$ は企業の雇用者総数、$union$ は組合があれば 1、なければ 0 の変数、$grant$ は州政府からの職業訓練助成金を受けていれば 1、受けてなければ 0 の変数である。

データは企業の行う職業訓練が企業の生産性に寄与しているか、そして州政府の助成金はそれに貢献しているかというプログラム評価を行うためのものである。ここでは、一人当たりの訓練時間が雇用者総数、組合ダミー、助成金ダミーでどれぐらい説明できるかを見ている。

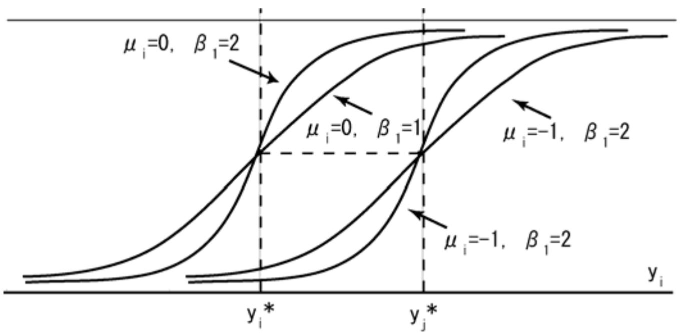
```
/*Panel Random Effect Tobit Model */
xttobit hrsemp employ union grant, i(fcode) ll(0)
xttobit lhrsemp lemploy union grant, i(fcode) ll(0)
```

結果は図表 5.6 に載せてある。雇用者総数はほとんど影響を与えていない。これを規模の指標と見れば、規模は職業訓練にはあまり関係がないと言えそうである。組合ダミーは負の効果を与えており、組合は職業訓練には消極的であることが示唆されている。州政府からの助成金は明らかに有意に正の効果を持っている。職業訓練は助成金を受けようが受けまいが企業の意思によって行っているものではあるが、職業訓練のための助成金が訓練時間に影響を与えるのは当然であるとも言える。Wooldridge (2003, p.445) で論じられているように、このプログラムでは助成申請した企業から順に助成を出しているということなので、職業訓練に熱心に取り組んでいる企業ほど助成を受け、その結果生産性が上がっていると言えそうである。その場合、助成金から訓練時間への因果関係とは逆になることに注意しなければならない。

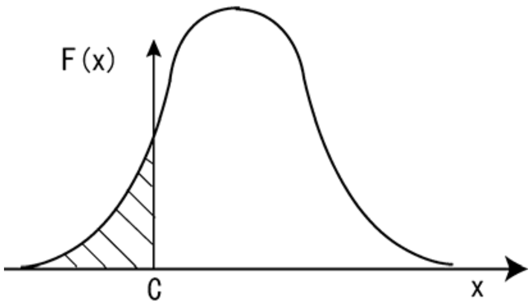
図表 5.1



図表 5.2



図表 5.3



図表 5.4

タイプ	尤度関数	被説明変数		
		y_1	y_2	y_3
1	$P(y_1<0)\cdot P(y_1)$	C	-	-
2	$P(y_1<0)\cdot P(y_1>0, y_2)$	B	C	-
3	$P(y_1<0)\cdot P(y_1, y_2)$	C	C	-
4	$P(y_1<0, y_3)\cdot P(y_1, y_2)$	C	C	C
5	$P(y_1<0, y_3)\cdot P(y_1>0, y_2)$	B	C	C

出典） Amemiya(1985, p.384), Table 10.2, Table 10.3
注1） C = **切断データ、** B = **二値（正・負）データ**

図表 5.5

Dependent Variable: enroll	Panel Logit		Panel Probit	
	Estimated Coefficient	z-statistics	Estimated Coefficient	z-statistics
educ	-0.1483	-3.80	-0.0695	-3.07
exper	-1.9365	-28.62	-1.0535	-30.34
expersq	0.1909	16.73	0.1069	18.27
black	-1.1899	-9.92	-0.6512	-9.96
_cons	1.4086	3.03	0.5911	2.21
Diagnostic Test				
Number of observation	12723		12723	
Number of groups	1881		1881	
Log Likelihood	-4473.4284		-4477.0404	
Wald test	chi2(4) = 1087.93 Prob>chi2 = 0.000		chi2(4) = 1236.71 Prob>chi2 = 0.000	
sigma_u	1.612		0.9016	
rho	0.4413		0.4484	
Likelihood ratio test for rho = 0	chibar2(01) = 271.84 Prob>chibar2 = 0.000		chibar2(01) = 288.80 Prob>chibar2 = 0.000	

図表 5.6

Dependent Variable: lhrsemp	Panel Tobit	
	Estimated Coefficient	z-statistics
lemploy	0.0455	0.35
union	-0.5253	-1.49
grant	2.6512	13.80
_cons	0.8064	1.67
Diagnostic Test		
Number of observation	390	
Number of groups	135	
Log Likelihood	-592.0544	
Wald test	chi2(3) = 192.78 Prob>chi2 = 0.000	
sigma_u	1.3889	
sigma_e	1.1841	
rho	0.5791	