Tobit モデル[1]

1 Tobit モデルの使いみち

被説明変数がある限られた範囲の値しか取らない状況，あるいは，なんらかの条件に当てはまったとき（当てはまらなかったとき）にはデータが観測できない状況では Tobit モデルが用いられる．このような被説明変数を**制限従属変数**（LDV: Limited Dependent Variable）とも呼ぶ．広義の Tobit モデルはさらに 2 つにわけて考えることができる．ひとつは，すべての観測値について説明変数も被説明変数も観測できるばあいである．このグループには，**端点解**（corner solution）のばあいや，トップコーディング等による**打ち切り**（censored）データのばあいが含まれる．いまひとつは，なんらかの条件に当てはまった観測値については被説明変数が観測できず，そのような観測値を推定に利用できないケースであり，**切断**（truncated）データと呼ばれる．このような状況はを**標本選択（サンプルセレクション）**（sample selection）とも呼ばれる．

データが打ち切られる，あるいは切断される観測できるための条件はあらかじめ分かっていることもあるし，分かっていないこともある．その条件が個々の観測値によって異なるばあいも，定数のばあいも，確率的に決まるばあいも考えられる．さらに，着目している推定式の説明変数にデータが観測できるかどうかが含まれるケース等も考えうる．広く Tobit モデルと呼ぶときには，これらの同時方程式体系をも含む[2]．このような Tobit モデルについても，離散選択モデルと同様，潜在変数（latent variable）を想定すると便利なことが多い．

2 Type I Tobit

被説明変数があらかじめ決められた範囲の値しか取らない状況は，標準的な打ち切り Tobit モデル（standard censored Tobit），あるいは **Type I Tobit** モデルと呼ばれる．この標準的な Tobit モデルが適用できる典型的な状況は，家計や企業の最適化行動の結果として観測される消費量や生産量が，ミクロ経済学でいうところの端点解になっているばあいである．労働時間や，通常の財の購入はマイナスの値を取らないから，観測される被説明変数の値の範囲は非負の実数に限定される．タバコや酒等の嗜好財の消費を考えてみると，消費量は連続変数として観測できるものの，消費量がゼロという世帯も一定数存在するだろう．

被説明変数 y_i に対して，潜在変数 y_i^* を考えよう．潜在変数は制約がないときの最適解を表す．これまでと同じく，潜在変数は説明変数と誤差項の 1 次関数で表されるとし，誤

[1] Wooldridge (2006) *Introductory Econometrics: A Modern Approach*, Ch.17 を参考にしている．また，Wooldridge (2002) Ch.16, 17 も参照せよ．式番号は Wooldridge (2006) による．

[2] Tobit モデルの種々のタイプについては，Amemiya (1985) を参照せよ．Type I から Type V Tobit という分類は雨宮健による．

差項は説明変数の条件付きで平均ゼロの正規分布に従うとしよう．つまり，

$$y_i^* = \mathbf{x}_i\beta + u_i, \quad u_i|\mathbf{x}_i \sim N(0, \sigma^2) \tag{17.18}$$

と表現できるとしよう．分散均一性（homoskedastisity）を仮定している．さて，観測される被説明変数の値は y_i はあらかじめ決められた範囲に限定されるが，ここでは正の値のみを取ることができるとする．すなわち，潜在変数が正の値をとったときにはその値がそのまま観測されるが，負の値となったときにはゼロとして観測される．y_i が取り得る範囲があらかじめ決まっていれば，同じようにして尤度を構成できる．ここでは，下限のみがあって，その下限がゼロのケースを扱っている．このとき，

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* \geq 0 \\ 0 & \text{if} \quad y_i^* < 0 \end{cases} \tag{17.19}$$

と書くことができる．潜在変数が負の値をとるときには被説明変数 y_i の値がゼロとなるので，標準正規分布の累積分布関数 $\Phi(.)$ を用いて

$$\begin{aligned} P(y_i = 0|\mathbf{x}_i) &= P(y_i^* < 0|\mathbf{x}_i) = P(\mathbf{x}_i\beta + u_i < 0|\mathbf{x}_i) \\ &= P(u_i < -\mathbf{x}_i\beta|\mathbf{x}_i) = P\left(\frac{u_i}{\sigma} < -\frac{\mathbf{x}_i\beta}{\sigma}\Big|\mathbf{x}_i\right) = \Phi\left(-\frac{\mathbf{x}_i\beta}{\sigma}\right) = 1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right) \end{aligned}$$

と変形できる．ここで，誤差項の分布が説明変数 $\mathbf{x}_i$ と独立という条件を使っていることに注意しよう．潜在変数が正の値を取るときにはそのまま観測されるから，<u>確率密度</u> は標準正規分布の確率密度関数 $\Phi(.)$ を用いて

$$Pr(y_i = \mathbf{x}_i\beta + u_i|\mathbf{x}_i) = Pr(u_i = y_i - \mathbf{x}_i\beta|\mathbf{x}_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right)$$

したがって，サンプルが母集団からの無作為抽出で，誤差項が独立に同一の分布に従っているとすれば，各観測値の対数尤度は，

$$\log L_i = \mathbf{1}(y_i = 0)\log\left[1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right] + \mathbf{1}(y_i > 0)\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right)\right] \tag{17.22}$$

と表現され，最大化すべき対数尤度はこの個別の対数尤度の和となる．推定されるべきパラメタは，潜在変数の 1 次関数の係数パラメタ β，誤差項の標準偏差 σ である．統計アプリケーションでの出力は，通常の最小 2 乗推定に類似する．

係数の解釈

標準的な Tobit モデルで推定される係数 β は，説明変数 $\mathbf{x}_i$ のうちの対応する変数が 1 単位増加するときの潜在変数 y_i^* の期待値の増分を表している．通常の最小 2 乗推定では，これがすなわち，被説明変数 y_i の期待値の増分であるが，Tobit モデルではそうではない．これは，潜在変数が負の値を取るばあいには，被説明変数がゼロに「水増し」されてしまうためである．そこで，被説明変数の期待値の増分はどのように表されるのか考えてみよう．

まず，説明変数 $\mathbf{x}_i$ で条件付けたときの被説明変数の期待値 $E[y_i|\mathbf{x}_i]$ を求めよう[3]．被説明変数がとる値は場合分けされているから，この場合分けにしたがって，とりうる値とその値になる確率を計算する．つまり，

$$E[y_i|\mathbf{x}_i] = P(y_i = 0)E[y_i|y_i = 0, \mathbf{x}_i] + P(y_i > 0)E[y_i|y_i > 0, \mathbf{x}_i]$$

を計算すればよい．左辺第 1 項は，$y_i = 0$ のときのことだからゼロなので，

$$E[y_i|\mathbf{x}_i] = P(y_i > 0)E[y_i|y_i > 0, \mathbf{x}_i]$$

である．上で述べたように，$P(y_i = 0) = 1 - \Phi(\mathbf{x}_i\beta/\sigma)$ だから，

$$P(y_i > 0) = \Phi(\mathbf{x}_i\beta/\sigma)$$

である．次に，$E[y_i|y_i > 0, \mathbf{x}_i]$ を求めよう．$y_i > 0$ のとき $y_i = y_i^*$ だから，

$$E[y_i|y_i > 0, \mathbf{x}_i] = E[y_i^*|y_i^* > 0, \mathbf{x}_i] = E[\mathbf{x}_i\beta + u_i|\mathbf{x}_i\beta + u_i > 0, \mathbf{x}_i]$$

期待値の線形性から，$\mathbf{x}_i\beta$ と σ を期待値の外に出して，

$$E[y_i|y_i > 0, \mathbf{x}_i] = \mathbf{x}_i\beta + E[u_i|u_i > -\mathbf{x}_i\beta, \mathbf{x}_i] = \mathbf{x}_i\beta + \sigma E\left[\frac{u_i}{\sigma}\middle|\frac{u_i}{\sigma} > -\frac{\mathbf{x}_i\beta}{\sigma}, \mathbf{x}_i\right]$$

ここで，u_i/σ は標準正規分布に従う．標準正規分布において，ある値 c より大きな値をとることが分かっているときの確率密度は，その値 c より大きい部分を膨らませればよいので，

$$\frac{\phi(z)}{1 - \Phi(c)}$$

となるから，$z(\phi(z)/1 - \Phi(c))$ を c から ∞ について積分して，

$$E[z|z > c] = \frac{\phi(c)}{1 - \Phi(c)}$$

である[4]．$z = u_i/\sigma$，$c = -\mathbf{x}_i\beta/\sigma$ と読み替えると，

$$E\left[\frac{u_i}{\sigma}\middle|\frac{u_i}{\sigma} > -\frac{\mathbf{x}_i\beta}{\sigma}, \mathbf{x}_i\right] = \frac{\phi(u_i/\sigma)}{1 - \Phi(-\mathbf{x}_i\beta/\sigma)} = \frac{\phi(u_i/\sigma)}{\Phi(\mathbf{x}_i\beta/\sigma)}$$

なので，

$$E[y_i|y_i > 0, \mathbf{x}_i] = \mathbf{x}_i\beta + \sigma\frac{\phi(u_i/\sigma)}{\Phi(\mathbf{x}_i\beta/\sigma)}$$

[3] もちろん潜在変数の期待値は $E[y_i^*|\mathbf{x}_i] = \mathbf{x}_i\beta$ である．

[4] 標準正規分布の密度関数については $\phi'(z) = -z\phi(z)$ が成り立つから，$z\phi(z)$ の原始関数は $-\phi(z)$ であり，

$$\int_c^\infty \frac{z\phi(z)}{1 - \Phi(c)}dz = \frac{1}{1 - \Phi(c)}[-\phi(\infty) - (-\phi(c))] = \frac{1}{1 - \Phi(c)}[\phi(c)]$$

となる．ここで，$\lambda(c) = \phi(c)/\Phi(c)$ と定義する．この比は**逆ミルズ比（inverse Mill's ratio）**と呼ばれる．

$$E[y_i|y_i > 0, \mathbf{x}_i] = \mathbf{x}_i\beta + \sigma\lambda(\mathbf{x}_i\beta/\sigma) \tag{17.24}$$

この期待値は条件付き期待値（conditional expectation）とも呼ばれ，右辺第 2 項は，潜在変数がマイナスになったときに被説明変数がゼロに「水増し」される効果を反映している．逆ミルズ比と σ の積は $\mathbf{x}_i$ に依存していることに注意しよう．これらをまとめると，

$$E[y_i|\mathbf{x}_i] = \Phi(\mathbf{x}_i\beta/\sigma)\left[\mathbf{x}_i\beta + \sigma\frac{\phi(u_i/\sigma)}{\Phi(\mathbf{x}_i\beta/\sigma)}\right] = \Phi(\mathbf{x}_i\beta/\sigma)\mathbf{x}_i\beta + \sigma\phi(\mathbf{x}_i\beta/\sigma) \tag{17.25}$$

となり，$E[y_i|\mathbf{x}_i]$ は説明変数の非線形関数となる．仮定によりこの期待値は正の値をとる．最小 2 乗推定では被説明変数の条件付き期待値は説明変数の線形関数であり，誤差項の分散の大きさは期待値に影響しなかったことを思い出そう．

さて，説明変数が連続とすれば，説明変数 x_k が 1 単位大きくなるときの期待値の増分を計算できる．最小 2 乗推定ではその値は β_k で与えられるが，標準的な Tobit では簡単な形では表現できない．期待値の変分は，積の微分の公式から

$$\frac{\partial}{\partial x_k}E[y_i|\mathbf{x}_i] = \frac{\partial P(y_i > 0)}{\partial x_k}E[y_i|y_i > 0, \mathbf{x}_i] + P(y_i > 0)\frac{\partial}{\partial x_k}E[y_i|y_i > 0, \mathbf{x}_i] \tag{17.28}$$

それぞれの項目について計算してみよう．$P(y_i > 0) = \Phi(\mathbf{x}_i\beta/\sigma)$ だから，

$$\frac{\partial P(y_i > 0)}{\partial x_k} = (\beta_k/\sigma)\phi(\mathbf{x}_i\beta/\sigma) \tag{17.29}$$

また，$E[y_i|y_i > 0, \mathbf{x}_i] = \mathbf{x}_i\beta + \sigma\lambda(\mathbf{x}_i\beta/\sigma)$ だから，

$$\frac{\partial}{\partial x_k}E[y_i|y_i > 0, \mathbf{x}_i] = \beta_k + \beta_k\frac{d\lambda}{dc}(\mathbf{x}_i\beta/\sigma)$$

ここで，$d\lambda/dc = -\lambda(c)[c + \lambda(c)]$ より，

$$\frac{\partial}{\partial x_k}E[y_i|y_i > 0, \mathbf{x}_i] = \beta_k\left[1 - \lambda(\mathbf{x}_i\beta/\sigma)[\mathbf{x}_i\beta/\sigma + \lambda(\mathbf{x}_i\beta/\sigma)]\right] \tag{17.26}$$

これらを代入すると，

$$\begin{aligned}\frac{\partial}{\partial x_k}E[y_i|\mathbf{x}_i] = {}& \frac{\beta_k}{\sigma}\phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\left(\mathbf{x}_i\beta + \sigma\lambda\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right)\\ &+\Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\beta_k\left[1 - \lambda\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\left[\frac{\mathbf{x}_i\beta}{\sigma} + \lambda\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right]\right]\end{aligned}$$

となるので，$\phi(c) = \Phi(c)\lambda(c)$ を使って整理すると，

$$\frac{\partial}{\partial x_k}E[y_i|\mathbf{x}_i] = \beta_k\Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right) \tag{17.30}$$

となる．つまり，標準的な Tobit モデルを適用するのが適切な状況で OLS 推定を行うと，その傾きは $\Phi(\mathbf{x}_i\beta/\sigma)$ のぶんだけ過大評価される．この限界効果も説明変数 $\mathbf{x}_i$ に依存するので，効果を平均で評価した $\beta_k\Phi(\bar{\mathbf{x}}\beta/\sigma)$ を報告するか，効果の平均 $\beta_k \times n^{-1}\sum_{i=1}^{n}\Phi(\mathbf{x}_i\beta/\sigma)$ が報告される．説明変数がダミー変数の場合は，その変数が 1 のときの期待値と 0 のときの期待値の差が報告されることが多い．

特定化の問題

ここで扱っている標準的な Tobit モデルは，観測値の i.i.d. を仮定して対数尤度を構成しているから，この仮定が成り立たなければ推定量は一致性を持たない．たとえば，分散不均一性は OLS では一致性の問題とならないが，Tobit では一致性を損なう要因となる．また，離散選択モデルのときに示したように，内生性の問題があれば一致性は失われる．とはいえ，実際に真のデータ生成過程（DGP: Data Generating Process）は知りえないのだから，一致性を持つための条件がどれくらい尤もらしいか，のほうが問題となろう．

標準的な Tobit モデルにおける潜在的な重要な仮定は，$y_i^* > 0$ という条件付きでの $y_i > 0$ となる確率が，同じ条件付きでの y_i の期待値と強く連関している，という点であろう．y_i がゼロ以上となる確率と，y_i の条件付き期待値が別の要因で決まっているということもありうる．たとえば，生命保険の購入量と年齢の関係を考えてみよう．若いうちは生命保険に入る傾向が強くないだろうから，年齢と $y_i > 0$ となる確率は正の関係になろう．他方で，保険を購入した人たちだけに限ってみれば，年老いているほうが保険の価値は小さくなるから保険の購入量は小さくなり，$E[y_i|y_i > 0, x_i]$ と年齢とは負の関係を持つだろう．このような関係は標準的な Tobit モデルでは取り扱うことができない．

被説明変数が $y_i > 0$ となる確率と，同じ条件付きでの y_i の期待値が，異なる要因で決まると想定するモデルは，hurdle モデルや，two-part モデルと呼ばれるモデルで考えることができる．これらも広い意味では Tobit モデルとみなされている．

3　打ち切りデータ

標準的な Tobit モデルが想定する状況とよく似た状況に，データの**打ち切り**（censored）が起きている状況がある．これは，被説明変数があらかじめ決められた範囲を外れた値をとるときに，被説明変数が観測できない状況である．ただし，説明変数は観測できていることに注意しよう．このようなことが起きる典型的な例のひとつは，**トップコーディング**（top coding）の状況である．個票を用いた消費額や所得額を被説明変数にする回帰分析を行うとき，回答率の向上や個人情報保護の観点から，非常に大きな値は「〜万円以上」と一括してコーディングされることがある．この閾値をたとえば 2000 万円とすると，観測される所得 y_i は，実際の所得 y_i^* に対して，$y_i = \min(y_i^*, 2000\text{ 万円})$ となる．このとき，観測される被説明変数の値の範囲は 2000 万円以下と解釈できる．

標準的な Tobit モデルと同じく，被説明変数 y_i に対して，潜在変数 y_i^* を考えよう．潜在変数はトップコーディングの例であれば実際の所得や消費量に対応する．潜在変数は説明変数と誤差項の 1 次関数で表されるとし，誤差項は説明変数の条件付きで平均ゼロの正規分布に従う，すなわち

$$y_i^* = \mathbf{x}_i\beta + u_i, \quad u_i|\mathbf{x}_i \sim N(0, \sigma^2) \tag{17.36}$$

とする．さて，観測される被説明変数の値は y_i はあらかじめ決められた範囲に限定されるが，c_i を閾値としてトップコーディングが行われているとしよう．すなわち，潜在変数が

c_i より小さい値をとったときにはその値がそのまま観測されるが，大きな値となったときには観測されない．ただし，c_i より大きいということは判別される．このようなケースを上からの打ち切り（censoring from above）もしくは right censoring と呼ぶ．y_i がある値より大きなときにのみ観測できる状況は，censoring from below，もしくは left censoring と呼ばれる．Left censoring でも同じようにして尤度を構成できる．さてこのばあい，

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* < c_i \\ c_i & \text{if} \quad y_i^* \geq c_i \end{cases} \tag{17.37}$$

と書くことができる．ここで，閾値 c_i は既知でさえあれば観測値ごとに異なっていてもよい．潜在変数が $y_i^* > c_i$ のときには被説明変数 y_i の値が c_i となるので，標準正規分布の累積分布関数 $\Phi(.)$ を用いて

$$\begin{aligned} P(y_i = c_i|\mathbf{x}_i) &= P(y_i^* > c_i|\mathbf{x}_i) = P(\mathbf{x}_i\beta + u_i > c_i|\mathbf{x}_i) \\ &= P(u_i > c_i - \mathbf{x}_i\beta|\mathbf{x}_i) = P\left(\frac{u_i}{\sigma} > \frac{c_i - \mathbf{x}_i\beta}{\sigma}\Big|\mathbf{x}_i\right) = 1 - \Phi\left(\frac{c_i - \mathbf{x}_i\beta}{\sigma}\right) \end{aligned}$$

と変形できる．ここで，誤差項の分布が説明変数 $\mathbf{x}_i$ と独立という条件を使っていることに注意しよう．標準的な Tobit モデルと同様に，サンプルが母集団からの無作為抽出で，誤差項が独立に同一の分布に従っているとすれば，各観測値の対数尤度は，

$$\log L_i = \mathbf{1}(y_i = c_i)\log\left[1 - \Phi\left(\frac{c_i - \mathbf{x}_i\beta}{\sigma}\right)\right] + \mathbf{1}(y_i < c_i)\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right)\right] \tag{17.39}$$

と表現され，最大化すべき対数尤度はこの個別の対数尤度の和となる．推定されるべきパラメタは，潜在変数の 1 次関数の係数パラメタ β，誤差項の標準偏差 σ である．

標準的な Tobit モデルと同じく係数 β は潜在変数 y_i^* の動きを示すパラメタである．しかし，打ち切りデータのばあいにはデータが観測できないのは主体の行動の結果ではなく，データ収集（data collection）の問題だから，期待値や限界効果は通常の最小 2 乗回帰と同じく，β で評価すればよく，期待値は説明変数の線形関数とみてよい．

打ち切りデータの 1 つの応用は，期間分析（duration analysis）である．期間分析は，事象（event）が起きるまでの期間の長さを被説明変数とする分析手法であり，しばしばデータの入手可能性から，事象がまだ起きていない観測値もサンプルに含まれる．そのようなばあい，事象が起きるまでの期間が観測できた期間よりも長いという情報しか使うことができず，打ち切りデータとみなすことができる．もっとも，期間分析は生存期間分析（survival analysis）の枠組みで検討されることも多い．

4　サンプル・セレクション・モデル

打ち切り（censored）と似たような状況に，データの切断（truncated）が起きている状況がある．切断データのばあい，一定の条件を満たさない観測値（データの一部）は観測されず，被説明変数の値の情報は利用できない．つまり母集団からランダムサンプリングした

サンプルのうちの一部分しか利用できないことになるので，**標本選択（sample selection）**が起きている，とも言う．

標本選択の典型例は，企業が申し出る（offer）賃金率の推定である．労働者は企業が offer する賃金率を受け入れて働くか，受け入れずに働かないか（take it or leave it）どちらかであるとすれば，観測される賃金率は受け入れられた賃金率だけであり，受け入れられなかった賃金率は観測されない．このようなとき，一致性のある推定量を得るにはどうしたらよいのだろうか．

標本選択は問題になるのか?

サンプルセレクションは必ずしも常に問題となる，つまり，最小 2 乗推定量の一致性をおびやかすというわけではない．たとえば，母集団から無作為抽出したサンプルからさらに無作為抽出してサンプルを作れば，実質的には小さな無作為抽出標本を作ったことになるので，最小 2 乗推定量は一致性を持つ．他方，標準的な Tobit モデルにおいて端点解をとる観測値を除去した標本を用いれば，(17.24) 式から容易に想像されるように，最小 2 乗推定量は一致性を失う．

ここで，選択変数（selection indicator）s_i を導入しよう．選択変数 s_i は，$(y_i, \mathbf{x}_i)$ の組合せが全て観測された観測値に対しては 1 の値をとる指標変数（indicator variable）であり，$s_i = 0$ であるような観測値 i は利用できない．母集団から無作為抽出された観測値について，被説明変数・説明変数・誤差項のあいだに，いつものような線形関係が成り立つとしよう．

$$y_i = \mathbf{x}_i\beta + u_i, \quad E[u_i|\mathbf{x}_i] = 0 \tag{17.42}$$

もし，s_i の値にかかわらず全ての i の y_i が観測できていたとすれば，その標本について最小 2 乗推定を行うことによって一致性のある推定量を得ることができる．標本選択が起きているときには $s_i = 1$ となる観測値しか利用できないが，それらに限っても直交条件が成り立っている，つまり，

$$E[u_i|\mathbf{x}_i, s_i = 1] = 0$$

が成り立っていれば，最小 2 乗推定推定量は一致性をもつ[5]．この条件が成り立つような状況はおもに 3 通り考えることができる．第 1 は，選択変数 s_i が説明変数 $\mathbf{x}_i$ の関数となっているばあい，つまり標本選択が説明変数だけで決まっているばあいである．このとき，

[5]この条件は，次のように理解することもできる．いま，利用可能な観測値だけを使う回帰式を

$$s_i y_i = s_i \mathbf{x}_i \beta + s_i u_i \tag{17.44}$$

と書こう．利用できない観測値（$s_i = 0$）については全ての項がゼロになるので，β の推定のためになんの情報ももたらさない．さて，この回帰式を最小 2 乗推定して一致推定量を得るための条件は，誤差項と説明変数が直交する（内積がゼロ）だから，

$$E[(s_i\mathbf{x}_i)(s_iu_i)] = E[s_i\mathbf{x}_iu_i] = 0 \tag{17.45}$$

$$E[s_iu_i|s_i\mathbf{x}_i] = 0$$

が成り立てばよい．

$\{\mathbf{x}_i, s_i = 1\}$ が持つ情報は $\{\mathbf{x}_i\}$ が持つ情報と同じである．第 2 は，選択変数 s_i が説明変数 $\mathbf{x}_i$ とも誤差項 u_i とも独立であるばあいである．これは無作為抽出のケースに対応する．第 3 は，第 1 と第 2 のばあいの組合せで，選択変数 s_i が，説明変数 $\mathbf{x}_i$ と，誤差項と無相関の確率変数の関数となっているばあいである．いずれのばあいでも，選択変数は誤差項 u_i と無相関であることに注意しよう．

標本選択が最小 2 乗推定量の一致性を失わせるケースとして，打ち切りデータのうち打ち切られた観測値を利用しないばあいを考えてみよう．打ち切られるのは $u_i > c_i - \mathbf{x}_i\beta$ であるときだったから，選択変数は $s_i = \mathbf{1}(u_i < c_i - \mathbf{x}_i\beta)$ であり，誤差項 u_i と相関を持つ．したがって，このばあいには最小 2 乗推定量は一致性を持たない．

Heckman の 2 段階推定（Heckit）

サンプルセレクションに対処する基本的な方法のひとつが **Heckman の 2 段階推定（Heckit）**である．注目している回帰式はいつものような線形関係であり，

$$y_i = \mathbf{x}_i\beta + u_i, \quad E[u_i|\mathbf{x}_i] = 0 \tag{17.46}$$

とする．ここで，標本選択が起きていて，選択変数が

$$s_i = \mathbf{1}(\mathbf{z}_i\gamma + v_i \geq 0) \tag{17.47}$$

と表されるとする．$\mathbf{z}_i$ は標本選択を規定する変数群であり，外生変数とする．すなわち，注目している回帰式の誤差項 u_i と無相関であり，かつ，標本選択にまつわる誤差項 v_i とも無相関とする．Heckman の 2 段階推定が機能するためには，しばしば説明変数 $\mathbf{x}_i$ は $\mathbf{z}_i$ の部分集合である．つまり，全ての説明変数は $\mathbf{z}_i$ に含まれ，かつ，説明変数に含まれない外生変数が $\mathbf{z}_i$ に含まれている．のちに明らかになるように，$\mathbf{z}_i$ に含まれるが説明変数 $\mathbf{x}_i$ に含まれない変数は，2 段階最小 2 乗推定における除外された操作変数と同じ役割を果たす．これらの条件をまとめると，

$$E[u_i|\mathbf{x}_i, \mathbf{z}_i] = 0$$

である．選択変数にまつわる誤差項 v_i は変数 $\mathbf{z}_i$ と（したがって $\mathbf{x}_i$ と）独立であり，標準正規分布に従うと仮定する．誤差項 v_i は注目している回帰式の誤差項 u_i とは相関してもよく，その相関係数（のようなもの）を ρ とする．もし $\rho = 0$ であれば，標本選択によって最小 2 乗推定の一致性が失われない第 3 のケースに該当し，標本選択バイアスは発生しない．

さて，注目している回帰式（17.46）について，観測されるという条件付きの期待値をとることを考えよう．観測されるかどうかは $\mathbf{z}_i$ と v_i によって規定されるから，まず，これらについての条件付き期待値をとると，

$$E[y_i|\mathbf{z}_i, v_i] = E[\mathbf{x}_i\beta|\mathbf{z}_i, v_i] + E[u_i|\mathbf{z}_i, v_i]$$

説明変数 $\mathbf{x}_i$ は $\mathbf{z}_i$ の部分集合であり，また $\mathbf{z}_i$ は u_i と独立であると仮定しているから，

$$E[\mathbf{x}_i\beta|\mathbf{z}_i, v_i] = \mathbf{x}_i\beta, \quad E[u_i|\mathbf{z}_i, v_i] = E[u_i|v_i]$$

が成り立つから，

$$E[y_i|\mathbf{z}_i, v_i] = \mathbf{x}_i\beta + E[u_i|v_i]$$

となる．u_i と v_i が平均ゼロの 2 変量正規分布に従うとすると，パラメタ ρ について $E[u_i|v_i] = \rho v_i$ と書くことができ，

$$E[y_i|\mathbf{z}_i, v_i] = \mathbf{x}_i\beta + \rho v_i$$

さらに $s_i = 1$ について条件付き期待値をとると，

$$E[y_i|\mathbf{z}_i, s_i = 1] = \mathbf{x}_i\beta + \rho E[v_i|\mathbf{z}_i, s_i = 1]$$

右辺第 2 項は，$\mathbf{z}_i\gamma + v_i \geq 0$ のときの v_i の期待値だから逆ミルズ比となり，

$$E[y_i|\mathbf{z}_i, s_i = 1] = \mathbf{x}_i\beta + \rho\lambda(\mathbf{z}_i\gamma)$$

先に述べたように，もし $\rho = 0$ であれば右辺第 2 項が消えてしまい，標本選択の問題は発生しない．

さて，β を推定するためには，説明変数の項に加えて $\rho\lambda(\mathbf{z}_i\gamma)$ を追加して最小 2 乗推定すればよい．逆に言うと，標本選択の問題とは，$\lambda(\mathbf{z}_i\gamma)$ の項が省略変数（omitted variable）となる問題ともいえる．$\lambda(\mathbf{z}_i\gamma)$ は変数として観測されないが，v_i が正規分布に従うとすれば，probit 推定により，

$$P(s_i = 1|\mathbf{z}_i) = \Phi(\mathbf{z}_i\gamma) \tag{17.49}$$

を推定することができる．したがって，Heckman の 2 段階推定は以下のような手続きで行われる．

1. 全ての観測値を使って，被説明変数が観測できるかできないかの（s_i を被説明変数とする）probit 推定を行い，逆ミルズ比の推定値（$\hat{\lambda}_i = \lambda(\mathbf{z}_i\hat{\gamma})$）を計算する．

2. 被説明変数が観測される観測値を使って，y_i を $\mathbf{x}_i$ と $\hat{\lambda}_i$ に回帰した最小 2 乗推定を行う

標本選択によって最小 2 乗推定が一致性を失うかどうかは，2 段階目の推定での $\hat{\lambda}_i$ の係数 ρ がゼロであるかどうかを検定すれば分かる．もし $\rho = 0$ という帰無仮説が棄却できなければ，標本選択の問題は起きていないから，被説明変数が観測されたサンプルだけ用いて最小 2 乗推定を行っても一致性のある推定量が得られる．

Heckman の 2 段階推定の 2 段階目では被説明変数が観測される観測値を使って，

$$y_i = \mathbf{x}_i\beta + \rho\hat{\lambda}_i + e_i$$

を推定することになる．ここから，$\mathbf{z}_i$ について 2 つの注意すべき点が指摘できる．第 1 に，逆ミルズ比が一致推定されている必要があるから，$\mathbf{z}_i$ には全ての説明変数 $\mathbf{x}_i$ が含まれていたほうがよいということである．データが手元にあるのだから，全ての $\mathbf{x}_i$ を $\mathbf{z}_i$ に含ませることは難しくないだろうし，そうでなければ 1 段階目で省略変数バイアスをもたらすかもしれない．第 2 に，逆ミルズ比は $\mathbf{z}_i$ の非線形関数であるものの，しばしば線形に近いことから，$\mathbf{z}_i$ と $\mathbf{x}_i$ が同じ変数の組合せであれば，多重共線性の問題を発生しかねないということである．ちょうど，2 段階最小 2 乗法において除外された操作変数がなければ推定が行えないのと同じ状況である．それゆえ，選択変数についての 1 段階目の回帰に用いる $\mathbf{z}_i$ には，説明変数 $\mathbf{x}_i$ に含まれない外生変数が 1 つ以上，識別のために必要である．

Heckman の 2 段階推定は probit と最小 2 乗推定の組合せであるが，選択変数の誤差項 v_i と本来の誤差項 u_i が従う 2 変量正規分布の形状を規定できれば，直接に尤度関数を構成することもできる．Stata はデフォルトではこのような最尤推定を行う．

ここまで扱ってきたモデルは，**Type 2 Tobit モデル**とも呼ばれる．これは **Type 5 Tobit モデル**と呼ばれるより一般的なモデルの特殊形とみなすこともできる（Amemiya 1995）．

$$
\begin{aligned}
y_i &= \mathbf{x}_i\beta + u_i \\
& \quad \beta = \beta_1 \quad \text{if} \quad s_i = \mathbf{z}_i\gamma + v_i < 0 \\
& \quad \beta = \beta_2 \quad \text{if} \quad s_i = \mathbf{z}_i\gamma + v_i \geq 0
\end{aligned}
$$

このモデルでは，s_i の値によって β がスイッチするので，**switching regression** とも呼ばれる．あるいは，s_i によってレジームが選択されるとも解釈できるので **self-selection model** と呼ばれることもある．さらに，s_i が観察できないばあいにも拡張でき，そのようなばあいには**未知レジームのスイッチング回帰**と呼ばれる．

5 Stata code

標準的な Tobit モデルは，以下のようなコマンドラインで推定することができる．

```
tobit 被説明変数 説明変数, ll(#) ul(#)
```

ここで，`ll` は left-censoring limit，`rl` は right-censoring limit である．
Heckman の 2 段階推定法は，以下のようなコマンドラインで推定することができる．

```
heckman 被説明変数 説明変数, select( 選択変数 )
```

上の表現を使えば，説明変数は x_i，選択変数は z_i である．`heckman` コマンドは，デフォルトでは最尤法を用いた推定をするので，あえて Heckman の 2 段階推定を行う場合には，`twostep` オプションを，`select( )` のあとに加える必要がある．