

1 基礎的問題 (70点)

Latent variable が以下のような形で与えられる二値選択モデルについて考察する。

$$y^* = x\beta + u. \quad (1)$$

ただし、 $(1 \times k)$ の説明変数ベクトル x は確率ベクトル、誤差項 u は確率変数であるとする。Latent variable と観察される被説明関数の y の関係は以下のとおりである。

$$y = 1 \text{ if } y^* \geq 0, y = 0 \text{ if } y^* < 0. \quad (2)$$

1) ここで誤差項の条件つき分布が $u|x \sim N(0, 1)$ であるとする。このときの $E(y|x)$ を計算せよ。

2) 説明変数ベクトルの1番目の要素を x_1 として、その係数を β_1 とする。 x_1 が微小に変化したとき $E(y|x)$ がどれだけ変化するか限界効果を求めよ。この限界効果が一番大きくなるのはどのような x をもった個体かを限界効果の式から検討せよ。

3) 母集団から i.i.d. 標本 $\{y_i, x_i\}_{i=1}^n$ が得られるとしよう。上の問題と同じように誤差項が $u|x \sim N(0, 1)$ であるときに各観察値の log likelihood function l_i を求めよ。

4) 誤差項の条件つき分布が $u|x \sim N(0, 1)$ であるとき、条件なしの分布は $u \sim N(0, 1)$ に従う。しかしながら逆は真ではない。なぜかを説明せよ。

5) 説明変数 x が内生であるとは $u|x \sim N(0, 1)$ が成立しないことであるが、具体的に $u|x \sim N(x\gamma, 1)$ という形で内生性が発生しているとしよう。このとき $E(y|x)$ を計算して、 x の係数が β とはならないことを示し、Probit 推定量が一致性を持たないことを示せ。

6) 誤差項が不均一分散を持つとして、その不均一分散構造が $u|x \sim N(0, \exp(x\delta))$ と定式化できるとしよう。なぜ分散の部分が \exp 関数で定義されているかもっともらしい理由を説明せよ。

7) 上記の問題6)のような不均一分散構造が想定できるとき、log likelihood function をどのように定式化すれば β と δ の一致推定は可能か。

2 応用問題 (30点)

発展途上国においてはしばしば労働市場が労働法の保護が与えられるフォーマルセクターとそのような保護が与えられないインフォーマルセクターに分断されていると考えられている。このような現実を反映してブラジルの労働力調査においては、各労働者がフォーマルセクターで働いているかインフォーマルセクターで働いているかがダミー変数で記録されている。ある研究者がフォーマルセクターとインフォーマルセクターそれぞれの市場における賃金決定の構造を推定しようとしている。彼女は各労働者がフォーマルセクターで働くか、インフォーマルセクターで働くかが内生的に決定される以下のようなモデルを推定しようとしている。

$$y_1 = x\beta_1 + u_1, \quad (3)$$

$$y_2 = x\beta_2 + u_2, \quad (4)$$

$$y_3^* = z\gamma + v, \quad (5)$$

$$y_3 = 1 \text{ if } y_3^* \geq 0, \quad y_3 = 0 \text{ if } y_3^* < 0. \quad (6)$$

ただし、 y_1 はフォーマルセクターでの賃金率の自然対数、 y_2 はインフォーマルセクターでの賃金率の自然対数、 y_3 はフォーマルセクターで働いている場合に1を、インフォーマルセクターで働いている場合に0をとるダミー変数である。説明変数ベクトル x は定数項、各個人の教育年数、潜在経験年数、その二乗をふくむベクトルで常に観察される。説明変数ベクトル z には x に加えて親がフォーマルセクターで働いていたかどうかのダミー変数が追加されている。各セクターにおける賃金率である y_1, y_2 は各個人に関してそれぞれ定義できるが、実際に観察されるのは $y_3 = 1$ のときには $y_1, y_3 = 0$ のときには y_2 だけである。誤差項に関しては、 (u_1, u_2, v) は z から独立と仮定し、 $v \sim N(0, 1)$ ならびに $E(u_1|v) = \rho_1 v$ と $E(u_2|v) = \rho_2 v$ を仮定する。

1) $E(y_1|x, y_3 = 1)$ と $E(y_2|x, y_3 = 0)$ をそれぞれ計算せよ。なおここでは $v \sim N(0, 1)$ のとき、 $E(v|v > c) = \frac{\phi(c)}{1 - \Phi(c)}$ と $E(v|v \leq c) = -\frac{\phi(c)}{\Phi(c)}$ が使えることを使ってかまわない。

2) 上記の計算結果を踏まえ、どのような推定を行えばすべてのパラメータを一致推定できるか、順序を追って説明せよ。ここで、 u_1 と v の間と u_2 と v の間には多変数正規性が仮定されていないため最尤推定はできないことに留意せよ。

3) 非線形性によらない識別を達成するために必要な条件を述べよ。