

質的変数モデル(1)

別所俊一郎

2006年7月7日

Today's attraction

- 質的変数 (Limited dependent variable) モデルとは
- Linear Probability Model
- Probit, logit model

質的変数モデルとは

- 質的な選択，あるいは有限個の選択の結果を表す変数
 - ローンの諾否 / 行く行かない / するしない / できるできない / 公立私立 / ...
 - 交通手段 / 金融手段 / ...
- 有限個の（数少ない）値（しばしば整数値）をとる
 - 典型的には 2 値変数（binary variable）
 - 3 個以上の値をとる場合，順序がある（ordered） / ない（multinomial）
- 質的な選択の結果がどのような要因で決まるか？
 - 重回帰分析が有効？
 - 被説明変数が 0 か 1 しかとらないときの回帰直線？

2 値変数

- Binary outcomes (二項選択モデル)
 - 進学 / 喫煙開始 / 対外援助の受容 / 就職 / ...
 - これらが被説明変数になった場合の回帰分析
- ここで扱うテーマ：住宅ローンの諾否と人種
 - データ：1990年に Boston-Fed が作った Boston HMDA データ
 - 住宅ローンの貸し出しに人種が関係しているか
 - 貸す貸さないは銀行員の判断で，返済能力の有無で判定 (?)
 - ひとつの重要な指標は，ローンの返済額と所得の比率 (PI 比)
 - Figure 9.1. 散布図

2 値変数への OLS

- 被説明変数が 2 値変数であっても OLS で推定を行うことは可能
 - 普通の OLS ととくにちがうところはない
 - PI 比が 0.3 のときの当てはめ値は 0.20
- 被説明変数が 2 値変数のときの当てはめ値の意味とは?
 - Population regression function は , 説明変数で条件付けられたときの Y の期待値 .
$$E[Y|X_1, \dots, X_k]$$
 - 2 値変数の場合 , Y の期待値は Y が 1 の値をとる確率
 - 当てはめ値は , 説明変数 X で条件付けられた , Y が 1 の値をとる確率 (の予測値)

$$E[Y|X_1, \dots, X_k] = \Pr(Y = 1|X_1, \dots, X_k)$$

Linear Probability Model

- 被説明変数が 2 値変数のときの重回帰モデル
- Population regression function は説明変数で条件付けられた, 被説明変数が 1 の値をとる確率
- 当てはめ値は, 被説明変数が 1 となる確率の予測値
- 係数は, 説明変数が 1 単位変化したときの, 被説明変数が 1 となる確率の変化分
- 通常の OLS と変わらないので, t 値・ F 値・信頼区間の形成等はそのまま
- ただし, 被説明変数が一直線上に並ぶことはありえないので, R^2 は使えない

Linear Probability Model

- Boston-HMDA への応用
 - PI 比が高いほどローンの申し込みを拒否されやすい
 - ローン決定要因は他にもあるだろうから , omitted variable bias の可能性
- 欠点 : 回帰直線が 0 より下に延びたり , 1 より上に延びたりする
 - Population regression function は確率を表すはず
 - 確率は 0 と 1 の間の数値しか取らない
 - ??
 - 非線形の定式化が必要か .

Probit and Logit regression

- 被説明変数が 2 値変数のときに用いられる非線形回帰モデル
- 期待値（予測される確率）が 0 と 1 の間に収まるように定式化
- 確率分布関数を利用
 - Probit は標準正規分布（ Φ ）を使用
 - Logit はロジスティック（logistic）分布を使用
- 多岐選択モデルへの拡張も .
 - Probit は順序モデルへ
 - Logit は順序のない多岐選択モデルへ

Probit 回帰

- 説明変数がひとつのときのモデル

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

- たとえば , $\beta_0 = -2, \beta_1 = 3, X = 0.4$ であれば ,
 $\Phi(-2 + 3 \times 0.4) = \Phi(-0.8) = 0.21$ となり , $Y = 1$ となる確率は 21 %
- $\beta_1 > 0$ であれば , X が大きくなれば $Y = 1$ の確率が高くなる
- $\beta_1 < 0$ であれば , X が大きくなれば $Y = 1$ の確率が低くなる
- ただし , これ以上の直接の係数の解釈は難しい
- X の条件付確率 $\Pr(Y = 1|X)$ や , その変化の大きさを計算して解釈する
- 説明変数がひとつなら図を書くのもよい : S 字型

Probit 回帰

- 説明変数が 2 つ以上のモデル

$$\Pr(Y = 1|X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

- たとえば , $\beta_0 = -1.6, \beta_1 = 2, \beta_2 = 0.5, X_1 = 0.4, X_2 = 1$ であれば , $\Phi(z) = \Phi(0.3) = 0.38$ となり , $Y = 1$ の確率は 38%

- 非線形なので , 説明変数が変化したときの期待値

$\Pr(Y = 1|X_1, X_2)$ の計算には注意

1. もとの値 X での当てはめ値を求める
2. 少し変化させた値 $X + \Delta X$ での当てはめ値を求める
3. 2 つの当てはめ値の差を計算する

- Boston-HMDA への応用

- PI 比が 0.3 0.4 では , 拒否確率は 0.097 0.159 [6.2%pts]

- PI 比が 0.4 0.5 では , 拒否確率は 0.159 0.239 [8.0%pts]

- PI 比 0.3 で白人 黒人では , 拒否確率は 0.075 0.233

Logit 回帰

- 標準正規分布ではなく，ロジスティック分布を用いる

$$\Pr(Y = 1|\mathbf{X}) = F(\mathbf{X}|\beta) = \frac{1}{1 + \exp(-\mathbf{X}\beta)}$$

- 計算が比較的容易なため，歴史的にはロジット回帰のほうがよく使われたことも
- Boston-HMDA への応用
 - PI 比 0.3 で白人 黒人では，拒否確率は 0.074 0.222
 - 得られる係数推定値は異なるが，示していることはほぼ似ている (Figure)

Probit, Logit, LPM

- 推定結果の比較
 - 得られる係数推定値は異なるが，示していることはほぼ似ている
 - 限界効果（marginal effect）などで比較
 - LPM でもそこそこ似た値を得る（adequate approximation）
- Probit と Logit は最尤法（ML 法）で推定する
 - 最尤法：Maximum Likelihood
 - 推定量は一致性を持ち，漸近的に正規分布に従う
 - t 値・F 値は同じように使える．信頼区間の形成も同様
 - Probit, Logit ていどであれば，統計ソフト上での使い方も同様

Boston-HMDA への応用

- ローンの諾否における人種差別の存在の統計的検証
- 説明変数候補 (Table 9.1)
 - 金銭的負担：PI 比，住宅支出-所得比
 - ローンの大きさ：住宅の価値に比べて
 - 信用履歴：消費者信用・過去の住宅ローン・公的な信用記録
 - 他の要因：住宅ローン保険の可否，自営 / 単身 / 高卒ダミー，分譲マンションダミー
 - 人種：14.2%が黒人
 - 住宅ローンの可否：12.0%が拒否される
- 2 値選択モデルの推定結果 (Table 9.2)
 - (1)(2)(3) は base specification，推定方法のみ異なる
 - (4)(5)(6) が感応度 (頑健性) チェック，説明変数の追加や非線形性

Boston-HMDA への応用

- Base specification
 - (1) は線形確率モデル
 - * PI 比が 0.1 増えると, 拒否確率は +4.5%pts
 - * 住宅価値に対してローンの比率が 95%以上なら, 拒否確率は +18.9%pts
 - * 公的信用記録が悪ければ, 拒否確率は +19.7%pts
 - * 住宅ローン保険が拒否されれば, 拒否確率は +70.2%pts
 - * 黒人なら拒否確率は, +8.4%pts
 - (2) はロジット, (3) はプロビット
 - * 黒人ダミーの効果を見るため, 他の説明変数は「平均値」を設定して, 当てはめ値の差をとる
 - * ロジットでは +6.0%pts, プロビットでは +7.1%pts

Boston-HMDA への応用

- Extended specification
 - (4) は学歴ダミーなどを追加：人種の効果は $+6.6\%$ pts
 - (5) は分譲マンションダミーを追加し，信用変数の非線形項を追加：人種の効果は $+6.3\%$ pts
 - (6) は交差項を追加：人種の効果は $+6.5\%$ pt
- 結論
 - 総じて人種の効果は統計的に有意に検出され， $+6.0 \sim 8.4\%$ pt
 - 他の説明変数の値を平均値におけば，黒人であることで拒否確率は 7.4% 14.5%

Boston-HMDA への応用

- 論点

- 内的妥当性：データの誤差，非線形性，交差項... 見直されたが同様の結果
- 内的妥当性：ローンの申込用紙に書かれないような金融情報の重要性．人種との相関
- 外的妥当性：1990 年の Boston にしかあてはまらないのでは？