QF5210A Financial Time Series

Chao ZHOU

Department of Mathematics, NUS

matzc@nus.edu.sg

4th Lecture, 29 June 2019

Outline

Limited dependent variable models

There are situations in finance where the dependent variable is qualitative.

The situation would be referred to as a limited dependent variable.

The term refers to any problem where the values that the dependent variables may take are limited to certain integers (e.g. 0, 1, 2, 3, 4) or even where it is a binary number (only 0 or 1).

For a clear and simple presentation, here we will concentrate on the case of binary dependent variables. For the more general multinomial case, see, e.g. Chirs Brooks's book 11.9

Linear probability model (LPM)

The LPM is the simplest way to deal with binary dependent variables (only 0 or 1).

Given a set of explanatory variables $\{x_{2i}, x_{3i}, \cdots, x_{ki}\}$, we consider the following model

$$p_i = \mathbb{P}(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i, \quad i = 1, \cdots, N \tag{1}$$

Then $\mathbb{P}(y_i = 0) = 1 - \mathbb{P}(y_i = 1) = 1 - p_i$.

Example

Modelling the probability of a firm i will pay a dividend $y_i = 1$ as a function of its market capitalisation x_{2i} (M\$)

The fitted model is

$$\widehat{p}_i = -0.3 + 0.012x_{2i} \tag{2}$$

where $\widehat{p}_i$ denotes the estimated probability for firm i.

The model suggests that for every $1M\$$ increase in capitalisation size, the probability that the firm will pay a dividend increases by 1.2%.

And a firm with a capitalisation of $50M\$$ will have $-0.3 + 0.012 \times 50 = 30\%$ probability of paying a dividend.

Limits of LPM

The LPM is simple and intuitive, however the above example show a severe problem.

For any firm with capitalisation less than 25M\$, the probability of dividend payment is negative ; for any firm with more that 88M\$, the probability is greater than one.

To overcome this limitation of the LPM, on can use the logit and probit model.

Logit model

The logistic function F is

$$F(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \tag{3}$$

where e is the exponential.

In fact, F is cumulative distribution function (CDF for short) of logistic distribution, thus taking values in $]0,1[$. So the logistic model estimated is

$$\begin{aligned} p_i = \mathbb{P}(y_i = 1) &= F(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i) \\ &= \frac{1}{1 + e^{-(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i)}} \end{aligned}$$

Estimation

We can define the inverse of the logistic function, denoted by G, the logit

$$G(p) = \ln \frac{p}{1-p} \tag{4}$$

Then $G(p_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i$.

Clearly, this model is not linear and thus is not estimable using OLS. MLE is usually used, which is that the parameters are chosen to jointly maximise a log-likelihood function (LLF for short).

The MLE can be found by standard software, e.g., the function *glm* in R.

MLE

Let $z_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki}$. For logit model, the likelihood function for each observation y_i will be

$$L_i = \left(\frac{1}{1+e^{-z_i}}\right)^{y_i} \left(\frac{1}{1+e^{z_i}}\right)^{1-y_i} \tag{5}$$

The likelihood function that we need will be based on the joint probability for all N observations. Assume that each observation on y_i is independent. Then the likelihood function will be given by

$$L(\theta) = \prod_{i=1}^{N} \left(\frac{1}{1+e^{-z_i}}\right)^{y_i} \left(\frac{1}{1+e^{z_i}}\right)^{1-y_i} \tag{6}$$

and the log-likelihood function which will be maximised

$$LLF(\theta) = -\sum_{i=1}^{N} \left(y_i \ln(1+e^{-z_i}) + (1-y_i)\ln(1+e^{z_i})\right) \tag{7}$$

Goodness of fit

The objective of ML is to maximise the value of the LLF, not to minimise the RSS.

Thus the the standard goodness of fit measures such as RSS, R^2 or $\bar{R}^2$ don't have any real meaning.

Two goodness of fit measures are commonly used for limited dependent variable models

(i) The percentage of y_i values correctly predicted,

$$\text{Percent correct predictions} = \frac{100}{N}\sum_{i=1}^{N}\left[y_i I(\hat{p}_i) + (1-y_i)(1-I(\hat{p}_i))\right] \tag{8}$$

where $I(\hat{p}_i) = 1$ if $\hat{p}_i > \bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$ and 0 otherwise. The higher this number, the better the fit of the model.

Kennedy (2003) suggests measuring goodness of fit as the percentage of $y_i = 1$ correctly predicted plus the percentage of $y_i = 0$ correctly predicted

$$\text{PCP} = 100 \times \left[\frac{\sum_{i=1}^{N} y_i I(\hat{p}_i)}{\sum_{i=1}^{N} y_i} + \frac{\sum_{i=1}^{N}(1-y_i)(1-I(\hat{p}_i))}{N - \sum_{i=1}^{N} y_i}\right] \tag{9}$$

(ii) Another measure is known as pseudo-R^2

$$\text{pseudo-}R^2 = 1 - \frac{LLF}{LLF_0} \tag{10}$$

where LLF is the maximised value of the log-likelihood function for the model and LLF_0 is the value of the log-likelihood function for a restricted model where all of the slope parameters are set to zero (i.e. the model contains only an intercept).

Since the likelihood value is between zero and one, the LLF must result in a negative number. As the model fit improves, pseudo-R^2 will rise.

Probit model

Instead of the CDF of logistic distribution, the CDF of normal distribution can be used

$$F(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} \tag{11}$$

This function provides a transformation to ensure that the fitted probabilities will lie between zero and one.

Estimation for the probit model will proceed in exactly the same way as for logit model, except that the form for the likelihood function will be slightly different.

For most applications, the logit and probit models will give very similar characterisations.

Both approaches are preferred to the LPM.

The only instance where the models may give non-negligibility different results occurs when the split of the y_i between 0 and 1 is very unbalanced – for example, when $y_i = 1$ occurs only 10% of the time.

In R, the function *polr* of the package MASS can be used for probit regression.

Latent variable model

It's possible to motivate the probit model as a latent variable model. Suppose there exists an auxiliary random variable

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i, \quad i = 1, \cdots, N \tag{12}$$

where $\epsilon \sim N(0, \sigma^2)$. Then y_i can be viewed as an indication for the sign of y_i^* :

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \text{ i.e. } -\epsilon_i < \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki}, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned} p_i = \mathbb{P}(y_i = 1) &= \mathbb{P}(y_i^* > 0) \\ &= \mathbb{P}(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i > 0) \\ &= \mathbb{P}(\epsilon_i > -(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki})) \\ &= \mathbb{P}(\epsilon_i < \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki}) \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki})^2}{2\sigma^2}} \\ &= F(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki}) \end{aligned}$$

Decomposition model

Rydberg and Shephard (2003) : modelling price change is to decompose it into three components and use conditional specifications for the components. At the ith trade

$$y_i \equiv P_{t_i} - P_{t_{i-1}} = A_i D_i S_i,$$

$$A_i = \begin{cases} 1, & \text{if a price changes ;} \\ 0, & \text{if no changes} \end{cases} \quad D_i|(A_i = 1) = \begin{cases} 1, & \text{if price} \uparrow \text{;} \\ -1, & \text{if } \downarrow \end{cases}$$

S_i is the size of the price change in ticks if there is a change at the ith trade and $S_i = 0$ if no price change. When there is a price change, S_i is a positive integer-valued random variable.
Let F_i be the information set available at the ith transaction.

$$\begin{aligned} \mathbb{P}(y_i|F_{i-1}) &= \mathbb{P}(A_i D_i S_i|F_{i-1}) \\ &= \mathbb{P}(S_i|D_i, A_i, F_{i-1})\mathbb{P}(D_i|A_i, F_{i-1})\mathbb{P}(A_i|F_{i-1}) \end{aligned}$$

Decomposition model specification

Model specification : A_i is a binary variable : $p_i = \mathbb{P}(A_i = 1|F_{i-1})$,

$$\mathrm{logit}(\mathrm{p_i}) = \ln(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}) = \mathbf{x}_i\boldsymbol{\beta} \text{ or } \mathrm{p_i} = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}.$$

$D_i|(A_i = 1)$ is also a binary variable : $\gamma_i = \mathbb{P}(D_i = 1|F_{i-1}, A_i = 1)$

$$\mathrm{logit}(\gamma_\mathrm{i}) = \ln(\frac{\gamma_\mathrm{i}}{1-\gamma_\mathrm{i}}) = \mathbf{z}_i\boldsymbol{\delta} \quad \text{or} \quad \gamma_\mathrm{i} = \frac{e^{\mathbf{z}_i\boldsymbol{\delta}}}{1+e^{\mathbf{z}_i\boldsymbol{\delta}}}$$

$S_i|(F_{i-1}, D_i = 1, A_i = 1) \sim 1 + g(\lambda_{u,i})$; $S_i|(F_{i-1}, D_i = -1, A_i = 1) \sim 1 + g(\lambda_{d,i})$, where $g(\lambda)$ is a Geometric distribution with parameter λ

$$\mathrm{logit}(\lambda_{\mathrm{j,i}}) = \ln(\frac{\lambda_{\mathrm{j,i}}}{1-\lambda_{\mathrm{j,i}}}) = \mathbf{w}_i\boldsymbol{\theta} \quad \text{or} \quad \lambda_{\mathrm{j,i}} = \frac{e^{\mathbf{w}_i\boldsymbol{\theta}}}{1+e^{\mathbf{w}_i\boldsymbol{\theta}}}, \quad \mathrm{j} = \mathrm{u}, \mathrm{d},$$

where $\mathbf{x}_i$, $\mathbf{z}_i$, $\mathbf{w}_i$ are finite-dimensional vectors consisting of elements of F_{i-1} and $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{\theta}$ are parameter vectors.

Decomposition model likelihood

Classify the ith trade, or transaction, into one of three categories
1. No price change : $A_i = 0$ with proba $(1 - p_i)$;
2. A price increase : $A_i = 1, \ D_i = 1$ with proba $p_i \gamma_i$. The size of the price increase follow $1 + g(\lambda_{u,i})$;
3. A price decrease : $A_i = 1, \ D_i = -1$ with proba $p_i(1 - \gamma_i)$. The size of the price decrease follow $1 + g(\lambda_{d,i})$.

$I_i(j) = 1$ if the jth category occurs, log likelihood function

$$
\begin{aligned}
& ln[\mathbb{P}(y_i|F_{i-1})] = I_i(1) ln[(1 - p_i)] \\
& + I_i(2)[ln(p_i) + ln(\gamma_i) + ln(\lambda_{u,i}) + (S_i - 1) ln(1 - \lambda_{u,i})] \\
& + I_i(3)[ln(p_i) + ln(1 - \gamma_i) + ln(\lambda_{d,i}) + (S_i - 1) ln(1 - \lambda_{d,i})]
\end{aligned}
$$

Decomposition model example

The overall log likelihood function is

$$ln[\mathbb{P}(y_1, ...y_n|F_0)] = \sum_{i=1}^{n} ln[\mathbb{P}(y_i|F_{i-1})],$$

which is a function of $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{\theta}_u$ and $\boldsymbol{\theta}_d$.

A simple ADS model : IBM data
Predictors : $\{A_{i-1},\ D_{i-1},\ S_{i-1}, V_{i-1}, x_{i-1}, BA_i\}$
V_{i-1} : volume of the previous trade (divided by 1000) ;
x_{i-1} : previous duration ;
BA_i : the prevailing bid-ask spread.

Decomposition model application

Model : Action : $\mathbb{P}(A_i = 1|F_{i-1}) = p_i$, $ln(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 A_{i-1}$;
Direction : $\mathbb{P}(D_i = 1|F_{i-1}, A_i = 1) = \gamma_i$, $ln(\frac{\gamma_i}{1-\gamma_i}) = \delta_0 + \delta_1 D_{i-1}$;
Size : $ln(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}) = \theta_{j,0} + \theta_{j,1} S_{i-1}$ with $j = d$ or u.
R : > m1=glm(Ai∼ Aim1, family=binomial)
> m2=glm(di∼dim1,family=binomial)

Result :

Paramater	β_0	β_1	δ_0	δ_1
Estimate	-1.057	0.0962	-0.067	-2.307
Std.Err.	0.011	0.018	0.017	0.036
Paramater	$\theta_{u,0}$	$\theta_{u,1}$	$\theta_{d,0}$	$\theta_{d,1}$
Estimate	2.235	-0.670	2.076	-0.506
Std.Err	0.066	0.034	0.043	0.017

Decomposition model implication

Implication :

(1.) $\mathbb{P}(A_i = 1 | A_{i-1} = 0) = exp(\beta_0)/(1 + exp(\beta_0)) = 0.258,$

$$\mathbb{P}(A_i = 1 | A_{i-1} = 1) = exp(\beta_0 + \beta_1)/(1 + exp(\beta_0 + \beta_1)) = 0.476.$$

(2.)

$$\mathbb{P}(D_i = 1 | F_{i-1}, A_i = 1) = \begin{cases} 0.483 & \text{if } D_{i-1} = 0, \quad A_{i-1} = 0 \\ 0.085 & \text{if } D_{i-1} = 1, \\ 0.904 & \text{if } D_{i-1} = -1, \end{cases}$$

(3.) Price increases :

$$S_i | (F_{i-1}, D_i = 1, A_i = 1) \sim 1 + g(\lambda_{u,i}); \ \mathrm{logit}(\lambda_{\mathrm{u,i}}) = 2.235 - 0.670 \mathrm{S}_{\mathrm{i-1}}.$$

Thank you for your attention.