ECONOMETRICS II, Fall 2025.

Bent E. Sørensen

Binary Choice (mainly follows Bruce Hansen's econometrics book).

Consider a variable Y with support $\{0, 1\}$. In econometrics, we typically call this class of models binary choice. Examples of binary dependent variables include: Purchase of a single item; Market entry; Participation; Approval of an application/patent/loan. The dependent variable maybe recorded as Yes/No, True/False, or 1/–1, but can always be written as 1/0. The goal in binary choice analysis is estimation of the conditional or response probability $P[Y = 1|X]$ given a set of regressors X. We may be interested in the response probability or some transformation such as its derivative—the marginal effect. A traditional approach to binary choice modeling (and limited dependent variable models in general) is parametric with estimation by maximum likelihood. There is also a substantial literature on semi-parametric estimation. In recent years, applied practice has tilted towards linear probability models estimated by least squares.

Let (Y, X) be random with $Y \in (0, 1)$ and $X \in R^k$. The response probability of Y

with respect to X is

$$P(x) = P[Y = 1|X = x] = E[Y|X = x].$$

The response probability completely describes the conditional distribution. The marginal effect, which is often the main object of interest, is

$$\frac{\partial}{\partial x}P(x) = \frac{\partial}{\partial x}P[Y = 1|X = x] = \frac{\partial}{\partial x}E[Y|X = x].$$

Models for the Response Probability

The most common models for the response probability $P(x)$ is the linear model, the Probit model, and the Logit model.

. *Linear Probability Model* : $P(x) = x'\beta$, where β is a coefficient vector. In this model, the response probability is a linear function of the regressors (although notice that you can have, say, a non-linear effect in age by including squared age as a regressor). The linear probability model has the advantage that it is simple to interpret because in the regression

$$Y = X'\beta + e\,,$$

the coefficient vector β captures the impact on predicted Y (that is, the probability of observing $Y = 1$) of a one unit increase in elements of X; i.e., the marginal effect.

Linear regression is consistent because

$$Y = P(X) + e\,,$$

with $E[e|X] = 0$, so it satisfies the conditions for OLS to be unbiased and BLUE (but here we are in a situation where we may not want to restrict ourself to linear estimators). Of course, the error e has the conditional distribution

$$e = \begin{cases} \{1 - P(X)\}, \text{ with probability } P(X) \\ \\ -P(X), \text{ with probability } 1 - P(X)\,. \end{cases} \tag{1}$$

The coefficients β measures the marginal effects (when X does not include nonlinear transformations, if you have a square term in X for example, you take the derivative w.r.t. X and report that for, say, the mean value of X). A disadvantage of the linear probability model is that it does not respect the $[0, 1]$ boundary. Fitted and predicted values from estimated linear probability models frequently violate these boundaries producing nonsense results in that dimension. Often, a researcher main focus is on the marginal effects, in particular whether a regressor affects the probabilities, in which case the linear model is fine. If you need the actual predicted probabilities, it is more problematic. The main reason that many (including me) uses the linear model is that the non-linear models

are inconsistent in panels when fixed effects (a dummy for each person, for example) is included (but we do not demonstrate that in this note).

Index Models: $P(x) = G(x'\beta)$ where $G(u)$ is a "link function" and β is a coefficient vector. This framework is also called a single index model where $x'\beta$ is a linear index function. In binary choice models, $G(u)$ is a distribution function which respects the probability bounds $0 \leq G(u) \leq 1$. In economic applications $G(u)$ is typically the normal or logistic distribution function, both of which are symmetric about zero so that $G(-u) = 1-G(u)$ (it may be more convenient to realize that this implies that for a random variable e with distribution function $G(e)$ the random variable $-e$ has the same distribution function $G(e)$—the main example here is the normal). We assume throughout this note that this symmetry condition holds.

Let $g(u)$ denote the density function of $G(u)$. In an index model,the marginal effect function is

$$\frac{\partial}{\partial x}P(x) = \beta g(x'\beta)\,.$$

Index models are only slightly more complicated than the linear probability model but have the advantage of respecting the $[0, 1]$ boundary. The two most common index models are the probit and logit.

Probit Model: $P(x) = \Phi(x'\beta)$, where $\Phi(u)$ is the standard normal distribution function. This is a traditional workhorse model for binary choice analysis. It is simple, easy to use, easy to interpret, and is based on the classical normal distribution.

Logit Model : $P(x) = \Lambda(x'\beta)$, where $\Lambda(u) = \frac{1}{1+exp(-u)}$ is the logistic distribution function. This is an alternative workhorse model for binary choice analysis. The logistic and normal distribution functions (appropriately scaled) have similar shapes so the probit and logit models typically produce similar estimates for the response probabilities and marginal effects. One advantage of the logit model is that the distribution function is available in closed form which speeds computation.

Linear Series Model: $P(x) = x_K\beta_K$ where $x_K = x_K(x)$ is a vector of transformations (typically a low-order polynomial) of x and β_K is a coefficient vector. A series expansion has the ability to approximate any continuous function including the response probability $P(x)$. The advantage of a linear series model is that its linear form allows the application of linear econometric methods. It is not guaranteed, however, to be boundary-respecting.

Latent Variable Interpretation

An index model can be interpreted as a latent variable model. Consider

$$Y^* = X'\beta + e$$

$$e \sim \ G(e)$$

$$Y = \mathbb{1}\{Y^* > 0\} = \begin{cases} 1 \text{ if } \ Y^* > 0 \\ \\ 0 \text{ if } \ Y^* \leq 0\,. \end{cases}$$

In this model, the observables are (Y, X) (but not Y^*). The variable Y^* is latent, linear in X and an error e, with the latter drawn from a symmetric distribution G. The observed binary variable Y equals 1 if the latent variable Y^* exceeds zero and equals 0 otherwise. The event $Y = 1$ is the same as $Y^* > 0$, which is the same as $X\beta + e > 0$. This means that the response probability is

$$P(x) = P[e > -x'\beta] = P[-e < x'\beta] = G(x'\beta)\,.$$

The final equality uses the assumption that $G(u)$ is symmetric about zero. This shows that the response probability is $P(x) = G(x'\beta)$, which is an index model with link function $G(u)$.

This latent variable model corresponds to a choice model, where Y^* is an individuals relative utility (or profit) of the options $Y = 1$ and $Y = 0$, and the individual selects the

option with the higher utility. We see that this structural choice model is identical to an index model with link function equalling the distribution of the error. It is a probit model if the error e is standard normal,and a logit model if e is logistically distributed. The error e is either standard normal or standard logistic, because the scale of the error distribution is not identifed. To see this, suppose that $e = \sigma u$ where u has a distribution $G(u)$ with unit variance. Then the response probability is

$$P[Y = 1|X = x] = P[\sigma u > -x'\beta] = P[u > -x'\beta^*],$$

where $\beta^* = \beta/\sigma$. Here β and σ are not separately identified—for example, if you double β, you could double σ and still have the same fit.

Likelihood

Probit and logit models are typically estimated by maximum likelihood. To construct the likelihood, we need the distribution of an individual observation. Recall that if Y is Bernoulli, such that $P[Y = 1] = p$ and $P[Y = 0] = 1 - p$, then Y has the probability mass function $p(y) = p^y (1 - p)^{(1-y)}, y = 0, 1$.

In the index model, $P[Y = 1|X] = G(X\beta)$, Y is conditionally Bernoulli, so its condi-

tional probability mass function is

$$\pi(Y|X) = G(X'\beta)^Y(1 - G(X'\beta)^{(1-Y)} = G(X'\beta)^Y G(-X'\beta)^{(1-Y)} = G(Z'\beta)\,,$$

where

$$Z = \begin{cases} X \text{ if } \ Y = 1 \\ \\ -X \text{ if } \ Y = 0\,. \end{cases} .$$

(Note, that many people would not define this Z variable, but instead have an "if" statement for whether Y is 0 or 1 in the summation below.) Taking logs and summing across observations, we obtain the log-likelihood function:

$$\ell_n(\beta) = \Sigma_{i=1}^n \log[G(Z_i'\beta)]\,.$$

For the probit and logit models this is

$$\ell_n^{\text{probit}}(\beta) = \Sigma_{i=1}^n \log[\Phi(Z_i'\beta)]$$

$$\ell_n^{\text{logit}}(\beta) = \Sigma_{i=1}^n \log[\Lambda(Z_i'\beta)]\,.$$

Define the first and (negative) second derivatives of the log distributionfunction: $h(x) = \frac{d}{dx}\log G(x)$ and $H(x) = -\frac{d^2}{dx^2}\log G(x)$. For the logit model, these equal

$$h^{logit}(x) = 1 - \Lambda(x)$$

$$H^{logit}(x) = \Lambda(x)(1 - \Lambda(x)),$$

and for the probit model

$$h^{probit}(x) = \frac{\phi(x)}{\Phi(x)}$$

$$H^{probit}(x) = \lambda(x)[x + \lambda(x)],$$

where the function $\lambda(x) = \frac{\phi(x)}{\Phi(x)}$ is known as the inverse Mills ratio. (Verify these claims yourself.)

Sometimes you have a structural model and you will derive a latent variable specification and estimate this. In other cases, you are less structural and use what discrete choice model is convenient. (In a panel with fixed effects, the non-linear models are not consistent, so the linear model is often used. I will return to this issue soon when we cover panel data models.).

Sometimes, you focus on whether a parameter is significant. Rightly so, but when you write an article, you have to provide economic interpretation. That is usually directly revealed by the coefficient in the linear model (unless you have polynomial terms or interactions) but for the Logit, Probit etc, you should usually show the marginal effect

evaluated at the mean of all the regressors for the variables of interest. Or depending on your context, you might show the marginal effect for young versus old. If you have dummies, say for gender, then you may show for example the partial effects evaluated at the mean of the non-gender variables for men and for women separately as well. Say, you estimate the likehood of owning stocks directly. Are women more likely than men (you could test that from the estimated coefficients), but also, is the probability of owning stock more sensitive to, say, income for men or for women. There are many other cases, but the point is in non-linear models, you have to spell out the economic interpretation of your estimates.