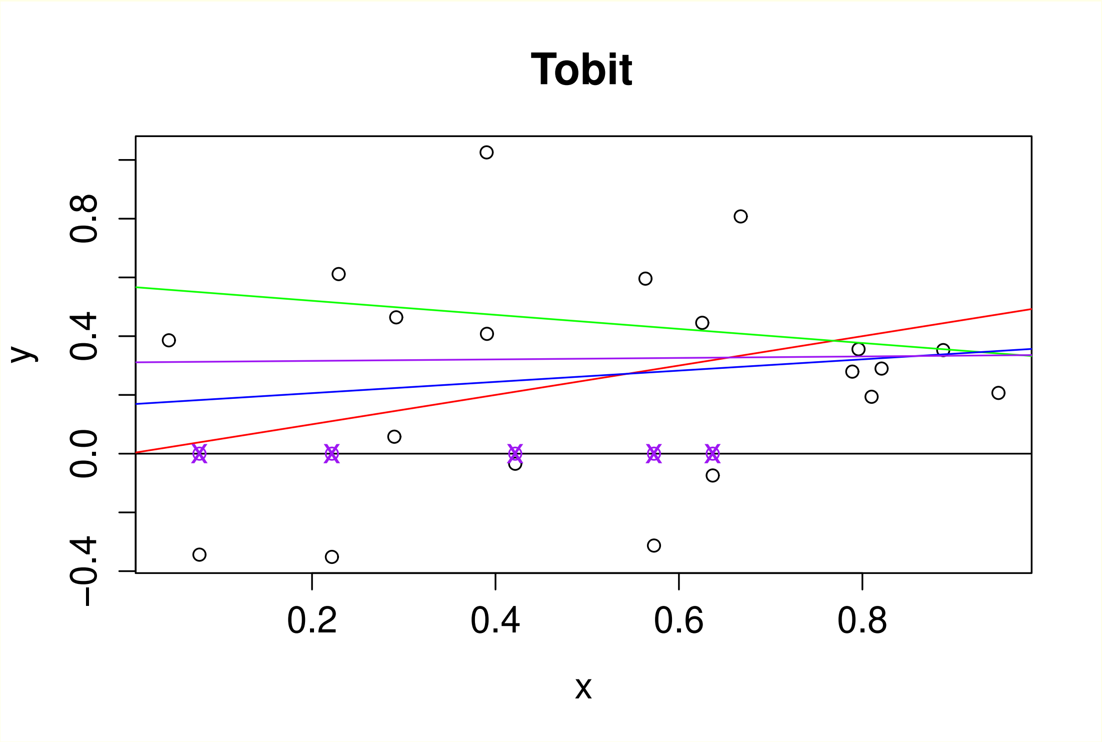
Limited dependent variables

February 20, 2020

Tobit

Why Tobin thought it up. Truncation vs. censoring.



Tobit

$$y_t = \max 0, X_t\beta + \varepsilon_t \, , \qquad \varepsilon_t \mid X_t \sim N(0, \sigma^2)$$

pdf for y_t is over lump of size 1 on 0 and Lebesgue measure on $(0, \infty)$. Thus it is

$$p(y_t \mid \beta) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{y_t - X_t\beta}{\sigma}\right) & \text{if } y_t > 0 \\ \Phi(-X_t\beta/\sigma) & \text{if } y_t = 0 \, . \end{cases}$$

Just maximize and/or plot the likelihood. (Could do data augmentation, generating unobserved $y_t < 0$ values in a Gibbs step.)

Censoring

$$y_t \mid x_t \sim \phi\left(\frac{y_t - x_t\beta)}{\sigma}\right)\left(1 - \Phi\Big(\frac{y_t - x_t\beta)}{\sigma}\Big)\right)^{-1}$$

That is, $y_| x_t$ is $N(x_t\beta, \sigma^2)$ *truncated* below at 0. The lower tail of the normal is missing, so to get the correct conditional distribution we have to scale up the normal pdf so it integrates to one over the truncated range.

For both cases

- Unlike SNLM, here the normality assumption matters, even for large samples. So we could also do Tobit-t, or consider even fancier (asymmetric, e.g.) pdf's for $\varepsilon_t \mid X_t$.

- The lower tail is identified asymptotically if $X_t\beta$ gets arbitrarily large, often. But even when it's not ID'd, tracing out how a prior on the distribution shape affects results is worthwhile.

LPM

- The linear robability model (LPM), logit, and probit all deal with dependent variables that take values in the two-element set $\{0, 1\}$.

- The LPM:

$$Y_t = X_t\beta + \varepsilon_t \,, \qquad E[\varepsilon_t \mid X_t] = 0 \,, \qquad P[Y_t = 1 \mid X_t] = X_t\beta \,.$$

- Can be estimated by OLS plus sandwich, so it's easy to use.

- It implies $\mathrm{Var}(\varepsilon_t \mid X_t) = X_t\beta(1 - X_t\beta)$

- This allows GLS, but is almost never used.

Problems with LPM

- Unless there are bounds on X, $X_t\beta$ can't be guaranteed to stay within the interval $[0, 1]$.

- Usually in fitted models, $X_t\hat{\beta}$ is outside $[0, 1]$ for only a very few observaions at most.

- The problem arises if you want to apply the model to predict Y for a new observation with an X that's not in the sample.

- It sometimes happens that for the sample at hand $X_t\beta$ remains well away from 0 and 1, say in the interval $[.2, .8]$. Then the model might be a good approximation, and even the use of GLS based on $\mathrm{Var}(Y_t - X_t\beta) = X_t\beta(1 - X_t\beta)$ could be useful.

Logit, Probit

- These are models of the form

$$P[Y_t = 1 \mid X_t] = f(X_t\beta) \, ,$$

where $f()$ is increasing monotonically and taking values in $[0, 1]$.

- Logit:

$$f(X_t\beta) = \frac{e^{X_t\beta}}{1 + e^{X_t\beta}} \, .$$

- Probit:

$$f(X_t\beta) = \Phi(X_t\beta) \, ,$$

where $\Phi()$ is the cdf of the standard normal distribution

Likelihood for LDV models

$$P[Y_t \mid X_t] = f(X_t\beta)^{Y_t}(1 - f(X_t\beta))^{1-Y_t}\,.$$

So assuming i.i.d. data,

$$P[Y_1, \ldots, Y_T \mid X_1, \ldots, X_T, \beta] = \prod_{t:Y_t=1} f(X_t\beta) \prod_{t:Y_t=0} (1 - f(X_t\beta))\,.$$

Measures of fit for LDV models

- Best is likelihood, or log likelihood.
- Apparently appealing alternatives can be based on counts of prediction errors of various types.
- Likelihood gives much more credit for an accurate prediction with a $f(X_t\beta)$ near one than for one with f near .51, and correspondingly more penalty for errors with high predicted probability of the opposite result.
- Example: Predicting rare events.