Nothing beyond lecture notes to will be covered on the final.

Tobit (censored regression) models.

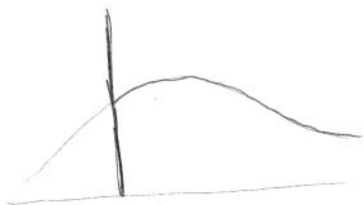
"What determines expenditure on a durable good?"
↳ a significant portion of people consume zero units of, say, refrigerators.

Latent variable $Y_i^* = X_i \beta_0 + \varepsilon_i$ where $Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > Y_0 \\ Y_0 & \text{else} \end{cases}$

$= Y_i^* 1\{Y_i^* > Y_0\} + Y_0 1\{Y_i^* \le Y_0\}$

$= \max\{Y_0, Y_i^*\}$

- We observe Y_i.



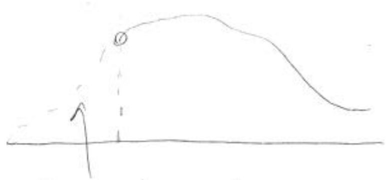
positive mass here! $\rightarrow Y_0 = 0$

- Censored regression model (Type I Tobit model)
  - Left censoring at zero
  - Censoring from below

If Y_0 is unknown this affects only the identification of the intercept. (We can only estimate the difference between the threshold and the intercept,

If, instead, $Y_i^* = \min\{Y_0, Y_i^*\}$, we say that we have right censored data.

Truncation



Just get rid of those observations

$Y_i^* = X_i \beta_0 + \varepsilon_i$

Observe $Y_i = X_i \beta_0 + \nu_i$ $\quad \nu_i = \varepsilon_i$ if $Y_i^* > 0$.

Even if $E[\varepsilon_i | X_i] = 0$ in censored regression model

$\Rightarrow E[Y_i^* | X_i] = X_i \beta_0$, but

$E[Y_i | X_i, Y_i > 0] \neq X_i \beta_0$
(truncated mean)

$\Leftrightarrow E[Y_i | X_i, \varepsilon_i > -X_i\beta_0] \neq X_i \beta_0$

Under censoring from below,
$E[Y_i | X_i, Y_i > 0] \geq X_i \beta_0$

Even $E[Y_i | X_i] \neq X_i \beta_0$.
(censored mean)

---

assumption: $\varepsilon_i | X_i \sim N(0, \sigma_0^2) \Rightarrow Y_i^* | X_i = (X_i\beta_0 + \varepsilon_i) | X_i \sim N(X_i\beta_0, \sigma_0^2)$

$$f(y_i | Y_i > 0) = \frac{d}{dy} \Pr[Y_i \leq y_i \mid Y_i \geq 0]$$

$$\Pr[Y_i \leq y_i \mid Y_i \geq 0] = \begin{cases} \dfrac{\Pr[Y_i \leq y_i \cap Y_i \geq 0]}{\Pr[Y_i \geq 0]} & \text{if } y_i \geq 0 \\ 0 & \text{else} \end{cases}$$

$$= \begin{cases} \dfrac{\int_0^{y_i} f(u)\,du}{\int_0^{\infty} f(u)\,du} & y_i \geq 0 \\ 0 & \text{else} \end{cases}$$

Fundamental Thm of Calculus

$$f(y_i | Y_i \geq 0) = \frac{f(y_i)}{\Pr[Y_i \geq 0]} = \frac{\frac{1}{\sigma_0} \varphi\left(\frac{y_i - x_i\beta_0}{\sigma_0}\right)}{\Pr[Y_i \geq 0]}$$

Since $\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(z-\mu)^2\right\} = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\}$
$= \frac{1}{\sigma} \varphi\left(\frac{z-\mu}{\sigma}\right)$.

also

$$\Pr[Y_i \geq 0] = \Pr[Y_i^* \geq 0] = \Pr[X_i\beta_0 + \varepsilon_i \geq 0]$$
$$= \Pr[\varepsilon_i \geq -X_i\beta_0] = \Pr\left[\frac{\varepsilon_i}{\sigma_0} \geq -\frac{X_i\beta_0}{\sigma_0}\right]$$
$$= 1 - \Phi\left(-\frac{X_i\beta_0}{\sigma_0}\right) = \Phi\left(\frac{X_i\beta_0}{\sigma_0}\right)$$

Thus, $f(y_i | Y_i \geq 0) = \dfrac{\frac{1}{\sigma_0}\varphi\left(\frac{y_i - X_i\beta_0}{\sigma_0}\right)}{\Phi\left(\frac{X_i\beta_0}{\sigma_0}\right)}$

The moments are still easy to calculate:

$$E[z | z \geq c] = \mu + \sigma \underbrace{\frac{\varphi\left(\frac{c-\mu}{\sigma}\right)}{1-\Phi\left(\frac{c-\mu}{\sigma}\right)}}$$

$$\equiv \lambda\left(\frac{c-\mu}{\sigma}\right)$$ inverse Mill's ratio (hazard function)

Thus, $E[Y_i | Y_i \geq 0, X_i] = X_i\beta_0 + \underbrace{\sigma_0 \lambda\left(\frac{-X_i\beta_0}{\sigma_0}\right)}_{\text{we call this the misspecification term}}$

since $\mu = X_i\beta_0$ and $c = 0$ here.

also, $V(z | z \geq c) = \sigma^2 \underbrace{\left[1 - \lambda\left(\frac{c-\mu}{\sigma}\right)\left[\lambda\left(\frac{c-\mu}{\sigma}\right) - \left(\frac{c-\mu}{\sigma}\right)\right]\right]}_{\text{can be shown to be } \leq 1}$

$$E[Y_i | X_i] = E[Y_i | Y_i \geq 0, X_i] \cdot Pr[y_i \geq 0 | X_i] + \underbrace{E[Y_i | Y_i < 0, X_i]}_{\substack{=0 \\ \text{since } Y_0 = 0}} \cdot Pr[y_i = 0 | X_i]$$

$$= \left[X_i\beta_0 + \sigma_0 \lambda\left(-\frac{X_i\beta_0}{\sigma_0}\right)\right] \Phi\left(\frac{X_i\beta_0}{\sigma_0}\right)$$

Typically, our objective is to estimate β_0.
Next time, we will consider estimation procedures.