14.387 Recitation 2
Probits, Logits, and 2SLS

Peter Hull

Fall 2014
Part 1: Probits, Logits, Tobits, and other Nonlinear CEFs
Going Latent (in Binary): Probits and Logits

Scalar bernoulli \( y_i \), vector \( x_i \). Assume

\[
y_i^* = x_i'\beta + v_i^*
\]
\[
y_i = 1 \{ y_i^* \geq 0 \}
\]

- \( y_i^* \) and \( v_i^* \): latent (unobserved) random variables
- What's the CEF (\( E[y_i|x_i] \))?
- Depends on the (conditional) CDF (of \( v_i^* \)):

\[
E[y_i|x_i] = P(y_i^* \geq 0|x_i)
\]
\[
= P(v_i^* \geq -x_i'\beta|x_i)
\]
\[
= 1 - F_{v^*}(x_i'\beta)
\]
\[
= F_{v^*}(-x_i'\beta)
\]

- Last line follows by CDF symmetry (usually assumed)
- Probit \( F_{v^*}() = ? \) Logit \( F_{v^*}() = ? \) Must the CEF actually be nonlinear?
Nonlinear Estimation

Two ways (at least) that $\beta$ is (probably) identified (where “probably” $\equiv$ “given some innocuous technical conditions”)

1. Maximum Likelihood (MLE):

$$
\beta^{MLE} = \arg\max_{\beta} \prod_i f_{y|x}(y_i|x_i, \beta) \\
= \arg\max_{\beta} \prod_i F_{\nu^*}(x_i^\prime \beta)^{y_i} (1 - F_{\nu^*}(x_i^\prime \beta))^{1-y_i}
$$

since $P(y_i = 1|x_i, \beta) = F_{\nu^*}(x_i^\prime \beta)$ and $P(y_i = 0|x_i, \beta) = 1 - F_{\nu^*}(x_i^\prime \beta)$

1. Nonlinear Least Squares (NLS)

$$
\beta^{NLS} = \arg\min_{\beta} E \left[ (y_i - F_{\nu^*}(x_i^\prime \beta))^2 \right]
$$

since $E[y_i|x_i] = F_{\nu^*}(x_i^\prime \beta) \implies y_i = F_{\nu^*}(x_i^\prime \beta) + \varepsilon_i$ with

$$
E[\varepsilon_i] = E[y_i - E[y_i|x_i]] = 0
$$

As with OLS, minimize expected squared prediction error, $E[\varepsilon^2_i]$
Maximum Likelihood

\[ \beta^{MLE} = \arg \max_{\beta} \prod_i F_{v^*}(x_i')^{y_i} (1 - F_{v^*}(x_i')^{\beta})^{1-y_i} \]

\[ = \arg \max_{\beta} \sum_i y_i \ln (F_{v^*}(x_i' \beta)) + (1 - y_i) \ln (1 - F_{v^*}(x_i' \beta)) \]

F.O.C.:

\[ 0 = \sum_i y_i \frac{f_{v^*}(x_i' \beta^{MLE})}{F_{v^*}(x_i' \beta^{MLE})} x_i - (1 - y_i) \frac{f_{v^*}(x_i' \beta^{MLE})}{1 - F_{v^*}(x_i' \beta^{MLE})} x_i \]

\[ = \sum_i \left( \frac{y_i}{F_{v^*}(x_i' \beta^{MLE})} - \frac{(1 - y_i)}{1 - F_{v^*}(x_i' \beta^{MLE})} \right) f_{v^*}(x_i' \beta^{MLE}) x_i \]

\[ = \sum_i \frac{(y_i - F_{v^*}(x_i' \beta^{MLE})) f_{v^*}(x_i' \beta^{MLE}) x_i}{F_{v^*}(x_i' \beta^{MLE})(1 - F_{v^*}(x_i' \beta^{MLE}))} \]

Plug-in estimator \( \hat{\beta}^{MLE} \) solves this in the sample
Nonlinear Least Squares

\[ \beta^{NLS} = \arg \min_{\beta} E \left[ (y_i - F_{v^*}(x_i' \beta))^2 \right] \]

F.O.C. (ignoring \(-2\) factor):

\[ 0 = E \left[ (y_i - F_{v^*}(x_i' \beta^{NLS})) f_{v^*}(x_i' \beta^{NLS}) x_i \right] \]

Plug-in estimator \( \hat{\beta}^{NLS} \) solves this in the sample

\[ 0 = \frac{1}{N} \sum_{i} (y_i - F_{v^*}(x_i' \hat{\beta}^{NLS})) f_{v^*}(x_i' \hat{\beta}^{NLS}) x_i \]

Look familiar?
MLE as Weighted Nonlinear Least Squares

Weighted NLS (like weighted least squares):

$$\beta^{\text{wNLS}} = \arg \min_\beta E \left[ W(x_i, y_i)(y_i - F_\nu^*(x_i'\beta))^2 \right]$$

for some (known) weight function $W(x_i, y_i)$. F.O.C.?

$$0 = \sum W(x_i, y_i)(y_i - F_\nu^*(x_i'\beta^{\text{wNLS}}))f_\nu^*(x_i'\beta^{\text{wNLS}})x_i$$

Recall

$$0 = \sum \frac{(y_i - F_\nu^*(x_i'\beta^{\text{MLE}}))f_\nu^*(x_i'\beta^{\text{MLE}})x_i}{F_\nu^*(x_i'\beta^{\text{MLE}})(1 - F_\nu^*(x_i'\beta^{\text{MLE}}))}$$

$\beta^{\text{MLE}}$ is a weighted NLLS estimator! But with what weights?
MLE as Weighted Nonlinear Least Squares (cont.)

\[ W^{MLE}(x_i, y_i) = \left( F_{v^*}(x_i \hat{\beta}^{MLE})(1 - F_{v^*}(x_i \hat{\beta}^{MLE})) \right)^{-1} \]

- MLE infeasible as one-step wNLS estimator (\( \hat{\beta}^{MLE} \) on both right and left of optimization)

- But recall another infeasible estimator

\[
\hat{\beta}^{GLS} = \arg\min_{\beta} \sum_i \left( \frac{y_i - x_i' \beta}{V_\varepsilon(x_i)} \right)^2
\]

where \( V_\varepsilon(x_i) \) is the conditional variance of \( \varepsilon_i \). (depends on \( \hat{\beta}^{GLS} \))

- We make GLS feasible by taking a first-step consistent estimate of \( V_\varepsilon(x_i) \) (by, say OLS), then solving

\[
\hat{\beta}^{FGLS} = \arg\min_{\beta} \sum_i \left( \frac{y_i - x_i' \beta}{V_\varepsilon(x_i)} \right)^2
\]
MLE as Weighted Nonlinear Least Squares (cont.)

\[ W^{MLE}(x_i, y_i) = \frac{1}{F_{\nu*}(x_i' \hat{\beta}^{MLE}) (1 - F_{\nu*}(x_i' \hat{\beta}^{MLE}))} = \frac{1}{\nu^*(x_i)} \]

Because \( y_i \) is bernoulli.

- Can take first-step consistent estimate of \( W^{MLE}(x_i, y_i) \) (by, say NLS) then solving wNLS FOC to get \( \hat{\beta}^{MLE_1} \)
- Use \( \hat{\beta}^{MLE_1} \) to get \( \hat{W}^{MLE_1} \rightarrow \hat{\beta}^{MLE_2} \rightarrow \hat{W}^{MLE_2} \rightarrow \ldots \)
- Iterating to convergence gives \( \hat{\beta}^{MLE} \)
Going Latent (with Truncation): Tobit

Assume

\[ y_i = \max(0, x_i' \beta + \varepsilon_i) \]
\[ \varepsilon_i \sim N(0, \sigma^2) \]

Useful normal fact: if \( w \sim N(\mu, \sigma^2) \) and \( c \) fixed,

\[ E[w|w > c] = \mu + \sigma \frac{\phi\left(\frac{\mu-c}{\sigma}\right)}{\Phi\left(\frac{\mu-c}{\sigma}\right)} \quad \text{and} \quad E[w|w < c] = \mu - \sigma \frac{\phi\left(\frac{c-\mu}{\sigma}\right)}{\Phi\left(\frac{c-\mu}{\sigma}\right)} \]

CEF:

\[ E[y_i|x_i] = E[y_i|x_i, y_i = 0]P(y_i = 0|x_i) + E[y_i|x_i, y_i > 0]P(y_i > 0|x_i) \]
\[ = (x_i' \beta + E[\varepsilon_i|x_i, \varepsilon_i > -x_i' \beta])P(\varepsilon_i > -x_i' \beta|x_i) \]
\[ = \left(x_i' \beta + \sigma \frac{\phi\left(x_i' \beta / \sigma\right)}{\Phi\left(x_i' \beta / \sigma\right)}\right) \Phi\left(x_i' \beta / \sigma\right) \]
\[ = x_i' \beta \Phi\left(x_i' \beta / \sigma\right) + \sigma \phi\left(x_i' \beta / \sigma\right) \]
Part 2: Some Facts about IV and 2SLS
Matrix-y IV

Setup:
- $n \times 1$ vector $Y$, $n \times r$ “endogenous” matrix $X_1$
- $n \times s$ matrix of “controls” $X_2$, $n \times t$ matrix of “instruments” $Z_1$

\[ X_1 = Z_1 \pi_1 + X_2 \pi_2 + \nu \]  
\[ Y = X_1 \beta_1 + X_2 \beta_2 + \epsilon \]

Terminology:
- (1) the first stage; (2) the second stage.
- Plugging (1) into (2) gives the reduced form:

\[ y = (Z \pi_1 + X_2 \pi_2 + \nu) \beta_1 + X_2 \beta_2 + \epsilon \]

\[ = Z_1 (\pi_1 \beta_1) + X_2 (\pi_2 \beta_1 + \beta_2) + (\nu \beta_1 + \epsilon) \]

- Model is identified if $t \geq r$ (just-identified if $t = r$)
- Exclusion restriction: $E[Z'\epsilon] = 0$ (weak), $E[\epsilon|Z] = 0$ (strong)
Matrix-y IV (cont.)

\[ X_1 = Z_1 \pi_1 + X_2 \pi_2 + \nu \]
\[ Y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon \]

Define:

\[ X \equiv \begin{bmatrix} X_1 & X_2 \end{bmatrix}, \quad n \times (r + s) \]
\[ Z \equiv \begin{bmatrix} Z_1 & X_2 \end{bmatrix}, \quad n \times (t + s) \]

Also define:

\[ P_Z \equiv Z(Z'Z)^{-1}Z', \quad P_2 \equiv X_2(X_2'X_2)^{-1}X_2, \quad M_2 \equiv I - P_2 \]

What’s \( P_Z Z = ? \) \( P_Z X = ? \) \( M_2 X_2 = ? \) \( P_Z P_Z = ? \) \( P_Z' = ? \)
2SLS is an IV Estimator

**IV Estimator:**

\[
\hat{\beta}^{IV} \equiv (W'X)^{-1} W'Y \\
W \equiv ZA
\]

where \( A = (t+s) \times (r+s) \) is some (possibly random) matrix.

Note that when we’re just-identified \( (t = r) \) \( A \) is (probably) invertible, so

\[
\hat{\beta}^{IV} \equiv (A'Z'X)^{-1} A'Z'Y \\
= (Z'X)^{-1} A^{-1} A'Z'Y = (Z'X)^{-1} Z'Y
\]

\( \implies \) all IV estimators are (numerically) equivalent when just-id

**Two-Stage Least Squares** sets \( A \equiv (Z'Z)^{-1} Z'X \). What’s \( W \)?
2SLS is a second-stage WLS/OLS regression

Two-Stage Least Squares is

\[
\hat{\beta}_{2SLS} = (\left((Z(Z'Z)^{-1}Z'X)'X\right)^{-1}(Z(Z'Z)^{-1}Z'X)'Y \\
= ((P_Z X)'X)^{-1}(P_Z X)'Y \\
= (X'P_Z X)^{-1}X'P_Z Y \\
= ((P_Z X)'P_Z X)^{-1}(P_Z X)'Y
\]

- (some kinda) Weighted Least Squares, by (3). What are the weights doing?
- (some kinda) Ordinary Least Squares, by (4). What are the regressors?
Just-ID IV is “reduced-form over first-stage”

\( \hat{\beta}^{2SLS} \) is OLS of \( Y \) on \( P_Z X \)

When \( r = 1 \) (one endogenous regressor), \( \hat{\beta}^{2SLS} \) is bivariate OLS of \( Y \) on \( M_2 P_Z X \)

\[
\hat{\beta}^{2SLS} \xrightarrow{p} \frac{\text{Cov}(y_i, \hat{x}_{1i}^*)}{\text{Var}(\hat{x}_{1i}^*)} = \frac{\text{Cov}(y_i, \hat{x}_{1i}^*)}{\text{Cov}(x_{1i}^*, \hat{x}_{1i}^*)}
\]

When \( t = r \) (just-identified),

\[
\text{Var}(\hat{x}_{1i}^*) = \pi_1^2 \text{Var}(Z_{1i}^*)
\]

\[
\text{Cov}(y_i, \hat{x}_{1i}^*) = \text{Cov}((\pi_1 \beta_1) Z_1 + X_2(\pi_2 \beta_1 + \beta_2) + (\nu \beta_1 + \epsilon), \pi Z_i^*)
\]

\[
= \pi_1^2 \beta \text{Var}(Z_{1i}^*)
\]

so that

\[
\hat{\beta}^{2SLS} \xrightarrow{p} \frac{\pi_1^2 \beta \sigma_{Z^*}^2}{\pi_1^2 \sigma_{Z^*}^2} = \left\{ \begin{array}{l}
\frac{\text{RF}}{FS}
\end{array} \right\} \frac{\pi_1 \beta}{\pi_1} = \beta
\]