EC708 Discussion 9
Limited Dependent Variable

Yan Liu

Department of Economics
Boston University

April 7, 2023

Outline

1. Estimation of Binary Response Models
2. Control Function
3. Semiparametric Binary Response Models

Estimation of Binary Response Models

MLE

Consider the binary response model

$$P(Y_t = 1|X_t) = F(X_t'\beta)$$

for some $\beta \in \mathbb{R}^k$ and $F : \mathbb{R} \to [0, 1]$. Once we choose a proper normalization, one can estimate β by MLE.

- Noting that $Y_t|X_t \sim \text{Bernoulli}(F(X_t'\beta))$, the conditional density is

$$L_t(\beta) = f(Y_t|X_t; \beta) = F(X_t'\beta)^{Y_t}(1 - F(X_t'\beta))^{1-Y_t}.$$

- The log-likelihood is

$$\bar{\ell}_T(\beta) = \sum_{t=1}^{T} \ln L_t(\beta) = \sum_{t=1}^{T} Y_t \ln F(X_t'\beta) + (1 - Y_t) \ln(1 - F(X_t'\beta)).$$

The score is

$$
\begin{aligned}
s_t(\beta) &= \frac{\partial \ln L_t(\beta)}{\partial \beta} \\
&= Y_t \frac{f(X_t'\beta)X_t}{F(X_t'\beta)} + (1 - Y_t)\frac{-f(X_t'\beta)X_t}{1 - F(X_t'\beta)} \\
&= \frac{f(X_t'\beta)X_t}{F(X_t'\beta)(1 - F(X_t'\beta))}[Y_t(1 - F(X_t'\beta)) - (1 - Y_t)F(X_t'\beta)] \\
&= \frac{Y_t - F(X_t'\beta)}{F(X_t'\beta)(1 - F(X_t'\beta))} f(X_t'\beta)X_t.
\end{aligned}
$$

Estimation of Binary Response Models

Hessian

Note that $Y_t^2 = Y_t$. The Hessian is

$$
\begin{aligned}
H_t(\beta) &= \frac{\partial s_t(\beta)}{\partial \beta'} \\
&= Y_t\Big[-\frac{f(X_t'\beta)^2}{F(X_t'\beta)^2} + \frac{f'(X_t'\beta)}{F(X_t'\beta)}\Big]X_tX_t' \\
&\quad - (1-Y_t)\Big[\frac{f(X_t'\beta)^2}{(1-F(X_t'\beta))^2} + \frac{f'(X_t'\beta)}{1-F(X_t'\beta)}\Big]X_tX_t' \\
&= -\frac{f(X_t'\beta)^2X_tX_t'}{F(X_t'\beta)^2(1-F(X_t'\beta))^2}[Y_t(1-F(X_t'\beta))^2 + (1-Y_t)F(X_t'\beta)^2] \\
&\quad + \frac{f'(X_t'\beta)X_tX_t'}{F(X_t'\beta)(1-F(X_t'\beta))}[Y_t(1-F(X_t'\beta)) - (1-Y_t)F(X_t'\beta)] \\
&= -\frac{(Y_t-F(X_t'\beta))^2}{F(X_t'\beta)^2(1-F(X_t'\beta))^2}f(X_t'\beta)^2X_tX_t' \\
&\quad + \frac{Y_t-F(X_t'\beta)}{F(X_t'\beta)(1-F(X_t'\beta))}f'(X_t'\beta)X_tX_t'.
\end{aligned}
$$

Estimation of Binary Response Models

Information Matrix Equality

Note that $E[Y_t - F(X_t'\beta)|X_t] = 0$ and

$$E[(Y_t - F(X_t'\beta))^2|X_t] = F(X_t'\beta)(1 - F(X_t'\beta)).$$

Hence,

$$\begin{aligned} E[s_t(\beta)s_t(\beta)'|X_t] &= \frac{E[(Y_t - F(X_t'\beta))^2|X_t]}{F(X_t'\beta)^2(1 - F(X_t'\beta))^2} f(X_t'\beta)^2 X_t X_t' \\ &= \frac{1}{F(X_t'\beta)(1 - F(X_t'\beta))} f(X_t'\beta)^2 X_t X_t'. \end{aligned}$$

Similarly,

$$E[H_t(\beta)|X_t] = -\frac{1}{F(X_t'\beta)(1 - F(X_t'\beta))} f(X_t'\beta)^2 X_t X_t'.$$

By the LIE, the information matrix equality holds:

$$I_0 = E[s_t(\beta)s_t(\beta)'] = -E[H_t(\beta)] = -H_0.$$

Control Function

Benchmark: Linear Models with Constant Coefficients

Consider the linear model:

$$Y_t = X_{1t}\beta_1 + X_{2t}'\beta_2 + U_t,$$

where X_{1t} is endogenous. Complete the model by adding the equation

$$X_{1t} = W_{1t}'\gamma_1 + X_{2t}'\gamma_2 + V_t.$$

Let $W_t = (W_{1t}', X_{2t}')'$ and $\gamma = (\gamma_1', \gamma_2')'$. Assume

- $E[W_t U_t] = 0$, $E[W_t V_t] = 0$;
- (no perfect multicollinearity) $\gamma_1 \neq 0$.

Correlation between U_t and V_t can be captured by

$$U_t = \lambda V_t + \eta_t, \quad E[V_t \eta_t] = 0,$$

where $\lambda = E[V_t U_t]/E[V_t^2]$. Then, $E[W_t \eta_t] = 0$ and

$$E[X_{1t}\eta_t] = E[W_t' \eta_t]\gamma + E[V_t \eta_t] = 0.$$

The linear model becomes

$$Y_t = X_{1t}\beta_1 + X_{2t}'\beta_2 + \lambda V_t + \eta_t.$$

Including V_t "controls for" endogeneity of X_{1t}. However, V_t is not observed.

Control Function

Benchmark: Linear Models with Constant Coefficients

Two-step control function procedure (linear reduced form):

1. Regress X_{1t} on W_t. Calculate the residual $\hat{V}_t$.
2. Regress Y_t on $X_t = (X_{1t}, X_{2t}')'$ and $\hat{V}_t$.

In this benchmark setting, the CF method

- is numerically identical to the regression (letting $\hat{X}_{1t} = X_{1t} - \hat{V}_t$)

$$\begin{aligned} Y_t &= (\hat{X}_{1t} + \hat{V}_t)\beta_1 + X_{2t}'\beta_2 + \lambda\hat{V}_t + \eta_t \\ &= \hat{X}_{1t}\beta_1 + X_{2t}'\beta_2 + (\beta_1 + \lambda)\hat{V}_t + \eta_t \Rightarrow \text{ 2SLS.} \end{aligned}$$

- produces a heteroskedasticity-robust Hausman test:
  simply test $H_0 : \lambda = 0$ (X_{1t} is exogenous) using the t-test

Control Function

Variation 1: Binary X_{1t}

When X_{1t} is binary, an alternative is to replace the linear reduced form with a binary response model:

$$\begin{aligned} Y_t &= X_{1t}\beta_1 + X_{2t}'\beta_2 + U_t, \\ X_{1t} &= 1\{W_{1t}'\gamma_1 + X_{2t}'\gamma_2 + V_t \geq 0\}. \end{aligned}$$

Assume that $(U_t, V_t) \perp W_t$, $V_t \sim N(0,1)$, and U_t is linearly related to V_t. Then, X_{1t} follows a probit model:

$$P(X_{1t} = 1 | W_t) = \Phi(W_t'\gamma).$$

Control Function

Variation 1: Binary X_{1t}

For some parameter λ,

$$E[Y_t|X_{1t}, W_t] = X_{1t}\beta_1 + X_{2t}'\beta_2 + \lambda\Big[\underbrace{X_{1t}\frac{\phi(W_t'\gamma)}{\Phi(W_t'\gamma)} - (1 - X_{1t})\frac{\phi(-W_t'\gamma)}{\Phi(-W_t'\gamma)}}_{\equiv r(X_{1t}, W_t'\gamma) \text{ "generalized error"}}\Big].$$

Two-step control function procedure (probit reduced form):

1. Estimate the probit model. Obtain the "generalized residuals":

$$\hat{r}_t \equiv X_{1t}\frac{\phi(W_t'\hat{\gamma})}{\Phi(W_t'\hat{\gamma})} - (1 - X_{1t})\frac{\phi(-W_t'\hat{\gamma})}{\Phi(-W_t'\hat{\gamma})}.$$

2. Regress Y_t on X_t and $\hat{r}_t$.

Control Function

Variation 2: Models Nonlinear in X_{1t}

Consider the model

$$Y_t = X_{1t}X_{2t}{}'\beta_1 + X_{2t}'\beta_2 + U_t$$

- IV approach treats each component in $X_{1t}X_{2t}$ as a separate endogenous variable.
- CF approach offers a parsimonious way to handle endogeneity: simply replace X_{1t} with $X_{1t}X_{2t}$ in the second-step regression.

Control Function

Variation 3: Correlated Random Coefficient Models

Consider the model

$$Y_t = X_{1t}\beta_{1t} + X_{2t}'\beta_2 + U_t,$$
$$X_{1t} = W_{1t}'\gamma_1 + X_{2t}'\gamma_2 + V_t.$$

Both β_{1t} and U_t might be correlated with X_{1t}. The object of interest is $\bar{\beta}_1 = E[\beta_{1t}]$. Write $\beta_{1t} = \bar{\beta}_1 + \tilde{\beta}_{1t}$, where $E[\tilde{\beta}_{1t}] = 0$. Then,

$$Y_t = X_{1t}\bar{\beta}_1 + X_{2t}'\beta_2 + \underbrace{X_{1t}\tilde{\beta}_{1t} + U_t}_{=\varepsilon_t}.$$

Control Function

Variation 3: Correlated Random Coefficient Models

Wooldridge (2003) shows that the 2SLS estimator is consistent under

- exogeneity: $E[U_t|W_t] = 0$, $E[\tilde{\beta}_{1t}|W_t] = 0$;
- constant conditional covariance assumption:

$$\mathrm{Cov}(X_{1t}, \tilde{\beta}_{1t}|W_t) = \mathrm{Cov}(X_{1t}, \tilde{\beta}_{1t}). \tag{1}$$

Card (2001) discusses situations where Condition (1) is likely to fail in simple models of schooling decisions.

Control Function

Variation 3: Correlated Random Coefficient Models

Control function approach: Assume

- U_t and $\tilde{\beta}_{1t}$ are linearly related to V_t:

$$E[U_t|V_t] = \lambda V_t, \quad E[\tilde{\beta}_{1t}|V_t] = \Psi V_t;$$

- $(U_t, \tilde{\beta}_{1t}, V_t) \perp W_t$.

The estimating equation is

$$E[Y_t|X_{1t}, W_t] = E[Y_t|X_{1t}, X_{2t}, V_t] = X_{1t}\bar{\beta}_1 + X_{2t}'\beta_2 + \lambda V_t + \Psi V_t X_{1t}.$$

Two-step procedure (Garen, 1984):

1. Regress X_{1t} on W_t. Calculate the residual $\hat{V}_t$.
2. Regress Y_t on X_t, $\hat{V}_t$, and $\hat{V}_t X_{1t}$.

Control Function

Variation 3: Correlated Random Coefficient Models

More flexibility:

- Allow any vector function $g(X_t)$ to have random slopes
  - $g(X_t)$ can include X_{1t}^2, $X_{1t}X_{2t}$, or higher-order polynomials and interactions (intercept is separated out)
  - CF regression: Y_t on 1, $g(X_t)$, $\hat{V}_t$, and $g(X_t)\hat{V}_t$.
- Allow $E[\tilde{\beta}_{1t}|V_t]$ to be nonlinear
  - E.g. $E[\tilde{\beta}_{1t}|V_t] = \Psi_1 V_t + \Psi_2(V_t^2 - \tau^2)$, where $\tau^2 = E[V_t^2]$.
  - CF regression: Y_t on X_t, $\hat{V}_t$, $\hat{V}_t X_{1t}$, $\hat{V}_t^2$, and $X_{1t} \cdot (\hat{V}_t^2 - \hat{\tau}^2)$, where $\hat{\tau}^2$ is the usual OLS variance estimate from the first stage.

Control Function

Nonlinear Models: Binary Choice

Consider the model

$$
\begin{aligned}
Y_t &= 1\{X_{1t}\beta_1 + X_{2t}'\beta_2 - U_t^* \geq 0\}, \\
X_{1t} &= W_{1t}'\gamma_1 + X_{2t}'\gamma_2 + V_t,
\end{aligned}
$$

where $(U_t^*, V_t) \perp W_t$ and $(U_t^*, V_t) \sim N(0, \Sigma)$ with $\Sigma = \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix}$.

By bivariate normality,

$$U_t^* = \lambda V_t + \eta_t \text{ for } \lambda = \rho/\sigma_v, \text{ where } \eta_t \perp V_t \text{ and } \eta_t \sim N(0, 1-\rho^2).$$

Putting together,

$$Y_t = 1\{X_{1t}\beta_1 + X_{2t}'\beta_2 - \lambda V_t - \eta_t \geq 0\},$$

where $\eta_t \perp (X_t, V_t)$.

Two-step control function procedure (Rivers and Vuong, 1988):

1. Regress X_{1t} on W_t. Calculate the residual $\hat{V}_t$.
2. Estimate a probit model including $\hat{V}_t$.

The average partial effects (APEs) are obtained by taking derivatives or changes of the average structural function (ASF):

$$\begin{aligned} \text{ASF}(x_1, x_2) &\equiv E_{U_t^*}[1\{x_1\beta_1 + x_2'\beta_2 - U_t^* \geq 0\}] \\ &= E_{V_t}[\Phi(x_1\beta_{1,\rho} + x_2'\beta_{2,\rho} - \lambda_\rho V_t)], \end{aligned}$$

where ρ subscript denotes division by $\sqrt{1-\rho^2}$. A consistent estimator is

$$\widehat{\text{ASF}}(x_1, x_2) = \frac{1}{T}\sum_{t=1}^{T} \Phi(x_1\hat{\beta}_{1,\rho} + x_2'\hat{\beta}_{2,\rho} - \hat{\lambda}_\rho \hat{V}_t).$$

Control Function

Nonlinear Models: Binary Choice

Blundell and Powell (2004) assume a full nonparametric structural model:

$$Y_t = g_1(X_{1t}, X_{2t}, U_t),$$
$$X_{1t} = g_2(W_{1t}, X_{2t}) + V_t,$$

where $(U_t, V_t) \perp W_t$ and $E[V_t] = 0$. Then,

- $g_2(W_t) = E[X_{1t}|W_t]$ is identified
- discrete X_{1t} is ruled out
- U_t depends on X_{1t} only through V_t:

$$U_t|X_{1t}, V_t \sim U_t|V_t$$

Control Function

Nonlinear Models: Binary Choice

By the LIE, the ASF is written as

$$\text{ASF}(x_1, x_2) \equiv E_{U_t}[g_1(x_1, x_2, U_t)] = E_{V_t}[E_{U_t|V_t}[g_1(x_1, x_2, U_t)]],$$

where

$$\begin{aligned} E_{U_t|V_t=v}[g_1(x_1, x_2, U_t)] &= E[g_1(x_1, x_2, U_t)|X_{1t} = x_1, X_{2t} = x_2, V_t = v] \\ &= E[Y_t|X_{1t} = x_1, X_{2t} = x_2, V_t = v] \\ &\equiv h(x_1, x_2, v) \text{ is identified.} \end{aligned}$$

A consistent estimator of the ASF is

$$\widehat{\text{ASF}}(x_1, x_2) = \frac{1}{T}\sum_{t=1}^{T} \hat{h}(x_1, x_2, \hat{V}_t), \text{ where } \hat{V}_t = X_{1t} - \hat{g}_2(W_t).$$

Semiparametric Binary Response Models

We consider the latent dependent variable model:

$$Y_t^* = X_t'\beta + U_t, \quad Y_t = 1\{Y_t^* > 0\},$$

where U_t has CDF G. If $U_t \perp X_t$, it induces the binary response model:

$$P(Y_t = 1|X_t) = G(X_t'\beta).$$

Without knowledge of G, the model is semiparametric with

- a finite-dimensional parameter of interest (β)
- an infinite-dimensional nuisance parameter ($G(\cdot)$)

If G is known, one can use the nonlinear least squares (NLS) estimator:

$$\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} (Y_t - G(X_t'\beta))^2.$$

Since G is unknown, Ichimura (1993) proposes to replace it with a kernel estimator $\hat{G}$ and solve

$$\min_{\beta} \frac{1}{T} \sum_{t=1}^{T} 1\{X_t \in A_x\}(Y_t - \hat{G}(X_t'\beta))^2.$$

The trimming term $1\{X_t \in A_x\}$ is introduced to guarantee that the density of $X_t'\beta$ is bounded away from 0 on A_x.

Semiparametric Binary Response Models

Manski (1975): Maximum Score Estimator

What if we further relax the assumption that $U_t \perp X_t$?

Manski (1975) proposes to maximize the predictive score function

$$S_T(\beta) = \sum_{t=1}^{T} Y_t 1\{X_t'\beta > 0\} + (1 - Y_t) 1\{X_t'\beta \leq 0\}.$$

- To ensure consistency, need the median of U_t given X_t to be zero.
- Can be interpreted as a least absolute deviations estimator:

$$\min_{\beta} \sum_{t=1}^{T} |Y_t - 1\{X_t'\beta > 0\}|.$$

Caveats:

- Nonnormal asymptotic distribution: median regression function is flat except at its discontinuity points
- Slow convergence rate: $T^{1/3}$ (Kim and Pollard, 1990)
- Does not allow estimation of the response probabilities and the APEs: unconditional distribution of U_t is not identified

Remedy: Smoothed Maximum Score Estimator (Horowitz, 1992)

Define a "smoothed" version of the predictive score function:

$$S_T^*(\beta) = \sum_{t=1}^{T} Y_t K(X_t'\beta/h_T) + (1 - Y_t)(1 - K(X_t'\beta/h_T)),$$

where K is analogous to a CDF, and $h_T \to 0$ as $T \to \infty$.

- The maximizer of $S_T^*(\beta)$ is asymptotically normal
- The convergence rate can be made arbitrarily close to $T^{1/2}$