Binary Models with Endogenous Explanatory Variables

Manuel Arellano

October 30, 2013

1. Introduction

- So far we considered linear and non-linear models with additive errors and endogenous explanatory variables.
- A simple case was a linear relationship between Y and X and an error U, where U was potentially correlated with X but not with an instrument Z:

$$Y = \alpha + \beta X + U, \qquad E(U) = 0, E(ZU) = 0. \tag{1}$$

  This setting was motivated in structural models.
- Now we wish to make similar considerations for binary index models of the form

$$Y = \mathbf{1}(\alpha + \beta X + U \geq 0).$$

- There are two important differences that we need to examine:

1. In the new models effects are heterogeneous, so that there is a difference between effects at the individual level and aggregate or average effects.
2. Instrumental variable techniques are no longer directly applicable because the model is not invertible and we lack an expression for U. So we have to consider alternative ways of addressing endogeneity concerns.

- To discuss these issues it is useful first to go back to the linear setting and re-examine endogeneity using an explicit notation for potential outcomes.

Potential outcomes notation

- When we are interested in (1) as a structural equation, we regard it as a conjectural relationship that produces potential outcomes for every possible value $x \in \mathcal{S}$:

$$Y(x) = \alpha + \beta x + U$$

- So we imagine that each unit has a value of U and hence a value of $Y(x)$ for each x.
- This is the way we think about structural models in economics, for example about a demand schedule that gives the conjectural demand $Y(x)$ for every possible price x.
- For each unit we only observe the actual value X that occurs in the distribution of the data, so that $Y = Y(X)$.
- If the assignment of values of X to units in the population is such that X and U are uncorrelated, β coincides with the regression coefficient of Y on X.
- If the assignment of values of Z to units in the population is such that Z and U are uncorrelated, β coincides with the IV coefficient of Y on X using Z as instrument.
- Consider two individuals with errors U and $U^{\dagger}$. Their potential outcomes differ:

$$\begin{aligned} Y(x) &= \alpha + \beta x + U \\ Y(x)^{\dagger} &= \alpha + \beta x + U^{\dagger} \end{aligned}$$

  but the effect of a change from x to x' will be the same for all individuals:

$$Y(x') - Y(x) = \beta(x' - x).$$

- In this sense we say that in models with additive errors the effects are homogeneous across units. We now turn to consider the situation in binary models.

Heterogeneous individual effects and aggregate effects

- Potential outcomes in the binary model are given by

$$Y(x) = \mathbf{1}(\alpha + \beta x + U \geq 0).$$

- The effect of a change from x to x' for an individual with error U is:

$$Y(x') - Y(x) = \mathbf{1}(\alpha + \beta x' + U \geq 0) - \mathbf{1}(\alpha + \beta x + U \geq 0).$$

- Suppose for the sake of the argument that $\beta > 0$ and $x' > x$. The possibilities are

value of U	$Y(x)$	$Y(x')$	$Y(x') - Y(x)$
$-U \leq \alpha + \beta x$	1	1	0
$-U > \alpha + \beta x'$	0	0	0
$\alpha + \beta x < -U \leq \alpha + \beta x'$	0	1	1

- Depending on the value of U the effects can be zero or unity, therefore they are heterogeneous across units.
- In these circumstances it is natural to consider an average effect:

$$\begin{aligned} E_U\left[Y(x') - Y(x)\right] &= E_U\left[\mathbf{1}(\alpha + \beta x' + U \geq 0)\right] - E_U\left[\mathbf{1}(\alpha + \beta x + U \geq 0)\right] \\ &= \Pr(-U \leq \alpha + \beta x') - \Pr(-U \leq \alpha + \beta x) \\ &= F(\alpha + \beta x') - F(\alpha + \beta x) \end{aligned}$$

  where F is the *cdf* of U.
- The average effect is simply the fraction of units in the population whose outcomes are affected by the change from x to x' (those with $\alpha + \beta x < -U \leq \alpha + \beta x'$).

Marginal effects

- If X is binary there is only one effect to consider. If X is continuous we can consider average marginal effects:

$$\frac{\partial E_U\left[Y\left(x\right)\right]}{\partial x}=\frac{\partial F\left(\alpha+\beta x\right)}{\partial x}=\beta f\left(\alpha+\beta x\right).$$

- Marginal effects can be regarded as a random variable associated with X.
- In this sense it may be of interest to obtain summary measures of its distribution, like the mean or the median. For example,

$$E_X\left[\beta f\left(\alpha+\beta X\right)\right].$$

Identification and estimation

- In models with additive errors, moment conditions of the form $E(ZU) = 0$ are often sufficient for identification, and GMM estimates can be easily constructed from their sample counterparts.
- In non-invertible models, GMM estimators are not directly available. In fact, the availability of instruments (a variable Z that is independent of U) by itself does not guarantee point identification in general.
- Next, we consider two specific models that fully specify the joint distribution of Y and X given Z.
- One is for a normally distributed X. It would have application to situations where X is a continuous variable.
- The other is for a binary X and leads to a multivariate probit model.

2. Binary outcome and continuous treatment

The normal endogenous explanatory variable probit model

- The model is

$$\begin{aligned} Y &= \mathbf{1}(\alpha + \beta X + U \geq 0) \\ X &= \pi' Z + \sigma_v V \end{aligned}$$

$$\begin{pmatrix} U \\ V \end{pmatrix} \mid Z \sim \mathcal{N}\left[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

- In this model X is an endogenous explanatory variable as long as $\rho \neq 0$.
- X is exogenous if $\rho = 0$.
- Joint normality of U and V implies that the conditional distribution of U given V is also normal as follows:

$$U \mid V, Z \sim \mathcal{N}\left(\rho V, 1 - \rho^2\right)$$

  or

$$\Pr(U \leq r \mid V, Z) = \Phi\left(\frac{r - \rho V}{\sqrt{1 - \rho^2}}\right).$$

  Therefore,

$$\Pr(Y = 1 \mid X, Z) = \Pr(\alpha + \beta X + U \geq 0 \mid V, Z) = \Phi\left(\frac{\alpha + \beta X + \rho V}{\sqrt{1 - \rho^2}}\right).$$

- Moreover, the density of $X \mid Z$ is just the normal linear regression density.

The normal endogenous explanatory variable probit model (continued)

- The joint probability distribution of Y and X given $Z = z$ is

$$f(y, x \mid z) = f(y \mid x, z) f(x \mid z)$$

or

$$\begin{aligned} \ln f(y, x \mid z) \quad &\propto \quad y \ln \Phi\left(\frac{\alpha + \beta x + \rho v}{\sqrt{1-\rho^2}}\right) + (1-y) \ln \left[1 - \Phi\left(\frac{\alpha + \beta x + \rho v}{\sqrt{1-\rho^2}}\right)\right] \\ &\quad -\frac{1}{2} \ln \sigma_v^2 - \frac{1}{2} v^2 \end{aligned}$$

where $v = (x - \pi' z) / \sigma_v$.

- Therefore, the log likelihood of a random sample of N observations conditioned on the z variables is $L\left(\alpha, \beta, \rho, \pi, \sigma_v^2\right) = L$:

$$\begin{aligned} L \quad = \quad & \sum_{i=1}^{N} \left\{ y_i \ln \Phi\left(\frac{\alpha + \beta x_i + \rho v_i}{\sqrt{1-\rho^2}}\right) + (1-y_i) \ln \left[1 - \Phi\left(\frac{\alpha + \beta x_i + \rho v_i}{\sqrt{1-\rho^2}}\right)\right] \right\} \\ & + \sum_{i=1}^{N} \left(-\frac{1}{2} \ln \sigma_v^2 - \frac{1}{2} v_i^2\right). \end{aligned}$$

- Note that under exogeneity ($\rho = 0$) this log likelihood function boils down to the sum of the ordinary probit and normal OLS log-likelihood functions:

$$L\left(\alpha, \beta, 0, \pi, \sigma_v^2\right) = L_{probit}(\alpha, \beta) + L_{OLS}\left(\pi, \sigma_v^2\right).$$

The control function approach
Two-step estimation of the normal model

- We can consider a two-step method:
- Step 1: Obtain OLS estimates $(\widehat{\pi}, \widehat{\sigma}_v)$ of the first stage equation and form the standardized residuals $\widehat{v}_i = \left(x_i - \widehat{\pi}' z_i\right) / \widehat{\sigma}_v$, $i = 1, ..., N$.
- Step 2: Do an ordinary probit of y on constant, x, and $\widehat{v}$ to obtain consistent estimates of $\left(\alpha^\dagger, \beta^\dagger, \rho^\dagger\right)$ where

$$\left(\alpha^\dagger, \beta^\dagger, \rho^\dagger\right) = \left(1 - \rho^2\right)^{-1/2} (\alpha, \beta, \rho) .$$

- Since there is a one-to-one mapping between the two, the original parameters can be recovered undoing the reparameterization.
- However, the fitted probabilities $\Phi\left(\widehat{\alpha}^\dagger + \widehat{\beta}^\dagger x_i + \widehat{\rho}^\dagger \widehat{v}_i\right)$ are in fact directly useful to get average derivative effects (more below).
- In general, two-step estimators are asymptotically inefficient relative to maximum likelihood estimation, but they may be computationally convenient.
- Ordinary probit standard errors calculated from the second step are inconsistent because estimated residuals are treated as if they were observations of the true first-stage errors.
- To get consistent standard errors, we need to take into account the additional uncertainty that results from using $(\widehat{\pi}, \widehat{\sigma}_v)$ as opposed to the truth.

Comparison with probit using fitted values

- Note that
$$Y = \mathbf{1}(\alpha + \beta X + U \geq 0) = \mathbf{1}\left(\alpha + \beta\left(\pi' Z\right) + \varepsilon \geq 0\right)$$
where $\varepsilon = U + \beta\sigma_v V$ is $\varepsilon \mid Z \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ with $\sigma_\varepsilon^2 = 1 + \beta^2\sigma_v^2 + 2\beta\sigma_v\rho$.
- If we run a probit of y on constant and $\widehat{x} = \widehat{\pi}' z$ we get consistent estimates of $\overline{\alpha} = \alpha/\sigma_\varepsilon$ and $\overline{\beta} = \beta/\sigma_\varepsilon$.
- Note that from estimates of $\overline{\alpha}$, $\overline{\beta}$, and σ_v we cannot back up estimates of α and β due to not knowing ρ.
- We cannot get average derivative effects either. So estimation of parameters of interest from this method (other than relative effects) is problematic.

The linear case: 2SLS as a control function estimator

- The 2SLS estimator for the linear IV model is $\widehat{\theta} = \left(\widehat{X}'\widehat{X}\right)^{-1} \widehat{X}'y$ where $\widehat{X} = Z\widehat{\Pi}'$ and $\widehat{\Pi} = X'Z\left(Z'Z\right)^{-1}$.
- The matrix of first-stage residuals is $\widehat{V} = X - Z\widehat{\Pi}'$. Typically, X and Z will have some columns in common. For those variables, the columns of $\widehat{V}$ will be identically zero. Let us call $\widehat{V}_1$ the subset of non-zero columns of $\widehat{V}$ (those corresponding to endogenous explanatory variables).
- It can be shown that 2SLS coincides with the estimated θ in the OLS regression

$$y = X\theta + \widehat{V}_1\gamma + \xi.$$

- Therefore, linear 2SLS can be regarded as a control function method.
- In the binary situation we obtained a similar estimator from a probit regression of y on X and first-stage residuals.
- An important difference between the two settings is that 2SLS is robust to misspecification of the first stage model whereas two-step probit is not.
- Examples of misspecifications occur if $E\left(X \mid Z\right)$ is nonlinear, if $Var\left(X \mid Z\right)$ is non-constant (heteroskedastic), or if $X \mid Z$ is non-normal.
- Another difference is that in the linear case the control-function approach and the fitted-value approach lead to the same estimator (2SLS) whereas this is not true for probit.

A semiparametric generalization

- Consider the model

$$\begin{aligned} Y &= \mathbf{1}(\alpha + \beta X + U \geq 0) \\ X &= \pi' Z + \sigma_v V \end{aligned}$$

  and assume that

$$U \mid X, V \sim U \mid V.$$

- In the previous parametric model we additionally assumed that $U \mid V$ was $\mathcal{N}(\rho V, 1 - \rho^2)$ and V was $\mathcal{N}(0, 1)$.
- The semiparametric generalization consists in leaving the distributions of $U \mid V$ and V unspecified.
- In this way

$$\Pr(Y = 1 \mid X, V) = \Pr(-U \leq \alpha + \beta X \mid X, V) = \Pr(-U \leq \alpha + \beta X \mid V)$$

  Thus

$$E(Y \mid X, V) = F(\alpha + \beta X, V)$$

  where $F(., V)$ is the conditional *cdf* of $-U$ given V.
- The function $F(., V)$ can be estimated nonparametrically using estimated first-stage residuals.
- This is a bivariate-index generalization of the semiparametric approaches to estimating single-index models with exogenous variables.

Constructing policy parameters

- To construct a policy parameter we need $p(x) = \Pr(-U \leq \alpha + \beta x)$. Note that

$$\Pr(-U \leq \alpha + \beta x) = \int \Pr(-U \leq \alpha + \beta x \mid v)\, dF_v \equiv E_V\left[F(\alpha + \beta x, V)\right].$$

- In the normal model

$$\Pr(-U \leq \alpha + \beta x) = \Phi(\alpha + \beta x)$$

- But this means that

$$\Phi(\alpha + \beta x) = E_V\left[\Phi\left(\frac{\alpha + \beta x + \rho V}{\sqrt{1-\rho^2}}\right)\right] \equiv E_V\left[\Phi\left(\alpha^\dagger + \beta^\dagger x + \rho^\dagger V\right)\right]. \qquad (2)$$

- A simple consistent estimate of $\Phi(\alpha + \beta x)$ is $\Phi\left(\widehat{\alpha} + \widehat{\beta} x\right)$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are consistent estimates. For example, using that $1 - \rho^2 = 1/\left(1 + \rho^{\dagger 2}\right)$, we may use

$$\left(\widehat{\alpha}, \widehat{\beta}\right) = \left(1 + \widehat{\rho}^{\dagger 2}\right)^{-1/2}\left(\widehat{\alpha}^\dagger, \widehat{\beta}^\dagger\right)$$

where $\left(\widehat{\alpha}^\dagger, \widehat{\beta}^\dagger, \widehat{\rho}^\dagger\right)$ are two-step control function estimates.

Constructing policy parameters (continued)

- Alternatively, using expression (2) a consistent estimate of $\Phi\left(\alpha+\beta x\right)$ in the normal model can be obtained as

$$\widehat{p}\left(x\right)=\frac{1}{N}\sum_{i=1}^{N}\Phi\left(\widehat{\alpha}^{\dagger}+\widehat{\beta}^{\dagger}x+\widehat{\rho}^{\dagger}\widehat{v}_{i}\right).$$

- In the semiparametric model this result generalizes to

$$\widetilde{p}\left(x\right)=\frac{1}{N}\sum_{i=1}^{N}\widehat{F}\left(\widetilde{\alpha}+\widetilde{\beta}x,\widehat{v}_{i}\right)$$

  where $\left(\widetilde{\alpha},\widetilde{\beta}\right)$ are semiparametric control function estimates and $\widehat{F}\left(.,.\right)$ is a non-parametric estimate of the conditional *cdf* of $-U$ given V.

3. Binary outcome and binary treatment

The endogenous dummy explanatory variable probit model

- The model is

$$\begin{aligned} Y &= \mathbf{1}(\alpha + \beta X + U \geq 0) \\ X &= \mathbf{1}(\pi' Z + V \geq 0) \end{aligned}$$

$$\begin{pmatrix} U \\ V \end{pmatrix} \mid Z \sim \mathcal{N}\left[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

- In this model X is an endogenous explanatory variable as long as $\rho \neq 0$. X is exogenous if $\rho = 0$.

The likelihood

- Let us introduce a notation for standard normal bivariate probabilities:

$$\Phi_2(r, s; \rho) = \Pr(U \leq r, V \leq s).$$

- The joint probability distribution of Y and X given Z consists of four terms:

$$\begin{aligned} p_{00} &= \Pr(Y = 0, X = 0) = \Pr(\alpha + \beta X + U < 0, X = 0) \\ &= \Pr\left(\alpha + U < 0, \pi' Z + V < 0\right) = \Phi_2\left(-\alpha, -\pi' Z; \rho\right) \end{aligned}$$

$$\begin{aligned} p_{01} &= \Pr(Y = 0, X = 1) = \Pr(\alpha + \beta + U < 0, X = 1) \\ &= \Pr(X = 1 \mid \alpha + \beta + U < 0) \Pr(\alpha + \beta + U < 0) \\ &= [1 - \Pr(X = 0 \mid \alpha + \beta + U < 0)] \Pr(\alpha + \beta + U < 0) \\ &= \Phi(-\alpha - \beta) - \Phi_2\left(-\alpha - \beta, -\pi' Z; \rho\right) \end{aligned}$$

$$\begin{aligned} p_{10} &= \Pr(Y = 1, X = 0) = \Pr(Y = 1 \mid X = 0) \Pr(X = 0) \\ &= [1 - \Pr(Y = 0 \mid X = 0)] \Pr(X = 0) = \Phi\left(-\pi' Z\right) - p_{00} \\ p_{11} &= 1 - p_{00} - p_{01} - p_{10}. \end{aligned}$$

Therefore, the log-likelihood is given by

$$L = \sum_{i=1}^{N} \left\{ (1 - y_i)(1 - x_i) \ln p_{00i} + (1 - y_i) x_i \ln p_{01i} + y_i (1 - x_i) \ln p_{10i} + y_i x_i \ln p_{11i} \right\}.$$

Average treatment effect

- In this model there are only two potential outcomes:

$$\begin{aligned} Y(1) &= \mathbf{1}(\alpha + \beta + U \geq 0) \\ Y(0) &= \mathbf{1}(\alpha + U \geq 0) \end{aligned}$$

- The average causal effect is given by

$$\theta = E[Y(1) - Y(0)] = \Phi(\alpha + \beta) - \Phi(\alpha).$$

- In less parametric specifications θ may not be point identified, but we may still be able to estimate average effects for certain sub-populations (more to follow).

Nonparametric binary model with binary endogenous regressor and instrument

- Consider the following model for $(0,1)$ binary observables (Y, X, Z):

$$\begin{aligned} Y &= \mathbf{1}(U_X \leq p_X) \\ X &= \mathbf{1}(V \leq q_Z) \end{aligned}$$

  where U_1, U_0 and V are uniformly distributed variates, independent of Z, such that (U_1, V) and (U_0, V) have copulas $C_1(u, v)$ and $C_0(u, v)$, respectively.

- Y is the dependent variable, X is the endogenous explanatory variable, and Z is the instrumental variable.
- Under exogeneity $C_1(u, v) = C_0(u, v) = uv$.
- A special case is the "switching" probit model

$$\begin{aligned} Y &= \mathbf{1}(\alpha + \beta X - U_X^* \geq 0) \\ X &= \mathbf{1}(\pi_0 + \pi_1 Z - V^* \geq 0) \end{aligned}$$

  where $p_X = \Phi(\alpha + \beta X)$, $U_X = \Phi(U_X^*)$, $q_Z = \Phi(\pi_0 + \pi_1 Z)$, $V = \Phi(V^*)$, and $C_1(u, v)$ and $C_0(u, v)$ are Gaussian copulas.

- A further specialization is the standard bivariate probit with endogeneity subject to the "monotonicity" constraint $U_1 \equiv U_0$.

Nonparametric binary model (continued)

- The data provides direct information about $\Pr(Y = j, X = k \mid Z = \ell)$ for $j, k, \ell = 0, 1$. Thus, given adding up constraints, there are 6 reduced form parameters.
- The structural parameters are p_0, p_1, q_0, q_1, $C_1(u, v)$ and $C_0(u, v)$.
- Because of the exogeneity of Z we have $q_\ell = \Pr(X = 1 \mid Z = \ell)$, so that q_0 and q_1 are reduced form quantities that are directly identified.
- The challenge is the identification of p_0 and p_1 or other probabilities associated with the potential outcomes.
- In the switching probit, the Gaussian copulas add just two extra structural parameters (the correlation parameters for the pairs (U_1, V) and (U_0, V)), so that the order condition for identification is satisfied with equality.
- In this model there are two potential outcomes:

$$\begin{aligned} Y_1 &= \mathbf{1}(U_1 \leq p_1) \\ Y_0 &= \mathbf{1}(U_0 \leq p_0) \end{aligned}$$

- The potential treatment indicators are:

$$\begin{aligned} X_1 &= \mathbf{1}(V \leq q_1) \\ X_0 &= \mathbf{1}(V \leq q_0). \end{aligned}$$

Identification

- The average treatment effect (ATE) is

$$\theta = E\left(Y_1 - Y_0\right) = p_1 - p_0.$$

- The mapping between reduced form and structural parameters is as follows. We observe q_0, q_1 and:

$$E\left(YX \mid Z=1\right) = C_1\left(p_1, q_1\right) \quad (3)$$
$$E\left(YX \mid Z=0\right) = C_1\left(p_1, q_0\right) \quad (4)$$

$$E\left[Y\left(1-X\right) \mid Z=1\right] = p_0 - C_0\left(p_0, q_1\right) \quad (5)$$
$$E\left[Y\left(1-X\right) \mid Z=0\right] = p_0 - C_0\left(p_0, q_0\right) \quad (6)$$

- If $C_1\left(u,v\right)$ and $C_0\left(u,v\right)$ are Gaussian copulas with correlation coefficients r_1 and r_0, it turns out that p_1 and r_1 are just identified from (3)-(4), whereas p_0 and r_0 are just identified from (5)-(6). Thus, the switching probit model is just identified.
- So normality is not testable in this model, it is just an identifying assumption. However, if $U_1 \equiv U_0$ bivariate probit places one over-identifying restriction.
- Alternative parametric copulas will produce different values of p_0 and p_1. So in general p_0 and p_1 are only set identified.

Identification (continued)

- To verify results (3)-(4) and (5)-(6) simply note that

$$E\left(YX \mid Z=1\right) = \Pr\left(Y=1, X=1 \mid Z=1\right) = \Pr\left(U_1 \leq p_1, V \leq q_1\right) = C_1\left(p_1, q_1\right)$$

$$E\left(YX \mid Z=0\right) = \Pr\left(Y=1, X=1 \mid Z=0\right) = \Pr\left(U_1 \leq p_1, V \leq q_0\right) = C_1\left(p_1, q_0\right)$$

$$E\left(1-X \mid Z=1\right) - E\left(1-X \mid Z=0\right) = q_0 - q_1,$$

$$\begin{aligned} E\left[Y\left(1-X\right) \mid Z=1\right] &= \Pr\left(Y=1, X=0 \mid Z=1\right) \\ &= \Pr\left(U_0 \leq p_0\right) - \Pr\left(U_0 \leq p_0, V \leq q_1\right) = p_0 - C_0\left(p_0, q_1\right) \end{aligned}$$

$$\begin{aligned} E\left[Y\left(1-X\right) \mid Z=0\right] &= \Pr\left(Y=1, X=0 \mid Z=0\right) \\ &= \Pr\left(U_0 \leq p_0\right) - \Pr\left(U_0 \leq p_0, V \leq q_0\right) = p_0 - C_0\left(p_0, q_0\right) \end{aligned}$$

LATE

- Suppose without lack of generality that $q_0 \leq q_1$. There are three subpopulations depending on an individual's value of V:
  - Always-takers: Units with $V \leq q_0$. They have $X_1 = 1$ and $X_0 = 1$. Their mass is q_0.
  - Compliers: Units with $q_0 < V \leq q_1$. Have $X_1 = 1$ and $X_0 = 0$. Their mass is $q_1 - q_0$.
  - Never-takers: Units with $V > q_1$. Have $X_1 = 0$ and $X_0 = 0$. Their mass is $1 - q_1$.
- Membership of these subpopulations is unobservable, but we observe their mass.
- The local ATE (LATE) is the ATE for the compliers:

$$\theta_{LATE} = E\left(Y_1 - Y_0 \mid q_0 < V \leq q_1\right).$$

- We have

$$E\left(Y_1 \mid q_0 < V \leq q_1\right) = \Pr\left(U_1 \leq p_1 \mid q_0 < V \leq q_1\right)$$
$$= \frac{\Pr\left(U_1 \leq p_1, V \leq q_1\right) - \Pr\left(U_1 \leq p_1, V \leq q_0\right)}{q_1 - q_0} = \frac{C_1\left(p_1, q_1\right) - C_1\left(p_1, q_0\right)}{q_1 - q_0}$$

  and similarly

$$E\left(Y_0 \mid q_0 < V \leq q_1\right) = \Pr\left(U_0 \leq p_0 \mid q_0 < V \leq q_1\right) = \frac{C_0\left(p_0, q_1\right) - C_0\left(p_0, q_0\right)}{q_1 - q_0}.$$

- Thus, LATE satisfies the difference in differences expression

$$\theta_{LATE} = \frac{\left[C_1\left(p_1, q_1\right) - C_1\left(p_1, q_0\right)\right] - \left[C_0\left(p_0, q_1\right) - C_0\left(p_0, q_0\right)\right]}{q_1 - q_0}$$

Link with instrumental variables

- Under monotonicity between X and Z (which the model assumes), θ_{LATE} coincides with the IV parameter (Imbens and Angrist 1994):

$$\theta_{LATE} = \frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(X \mid Z=1) - E(X \mid Z=0)}$$

- To see this, first note that using the previous results $E(Y_1 \mid q_0 < V \leq q_1)$ and $E(Y_0 \mid q_0 < V \leq q_1)$ are identified as:

$$E(Y_1 \mid q_0 < V \leq q_1) = \frac{E(YX \mid Z=1) - E(YX \mid Z=0)}{E(X \mid Z=1) - E(X \mid Z=0)}$$

$$E(Y_0 \mid q_0 < V \leq q_1) = \frac{E[Y(1-X) \mid Z=1] - E[Y(1-X) \mid Z=0]}{E(1-X \mid Z=1) - E(1-X \mid Z=0)}$$

- Next use the identities

$$E(Y \mid Z=1) = E(YX \mid Z=1) + E[Y(1-X) \mid Z=1]$$

$$E(Y \mid Z=0) = E(YX \mid Z=0) + E[Y(1-X) \mid Z=0]$$

$$E(X \mid Z=1) = E(X_1) = q_1, \qquad E(X \mid Z=0) = E(X_0) = q_0$$