1 A binary model with binary endogenous regressor and instrument

- Let us consider the following model for \((0,1)\) binary observables \((Y,D,Z)\):

\[
Y = 1(U_D \leq p_D)
\]
\[
D = 1(V \leq q_Z)
\]

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. In this model \(Y\) is the dependent variable, \(D\) is the endogenous explanatory variable, and \(Z\) is the instrumental variable.

- A special case is a switching probit model of the form

\[
Y = 1\left(\alpha + \beta D - \tilde{U}_D \geq 0\right)
\]
\[
D = 1\left(\pi_0 + \pi_1 Z - \tilde{V} \geq 0\right)
\]

where \(p_D = \Phi(\alpha + \beta D), U_D = \Phi(\tilde{U}_D), q_Z = \Phi(\pi_0 + \pi_1 Z), V = \Phi(\tilde{V})\), and \(C_1(u,v)\) and \(C_0(u,v)\) are Gaussian copulas. A further specialization is a standard bivariate probit with endogeneity subject to the “monotonicity” constraint \(U_1 \equiv U_0\).

- The data provides direct information about \(\Pr(Y = j, D = k \mid Z = \ell)\) 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 \(q_\ell = \Pr(D = 1 \mid Z = \ell)\), so that \(q_0\) and \(q_1\) are reduced form quantities and therefore always identifiable. The challenge is the identification of \(p_0\) and \(p_1\) or other probabilities associated with the potential outcomes.

- Note that in the switching probit model, the Gaussian copulas add just two extra structural parameters (i.e. the correlation coefficients of 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:

\[
Y_1 = 1(U_1 \leq p_1)
\]
\[
Y_0 = 1(U_0 \leq p_0)
\]
The potential treatment indicators are:

\[ D_1 = 1 (V \leq q_1) \]
\[ D_0 = 1 (V \leq q_0). \]

2 ATE, LATE and potential outcome distributions of compliers

• The average treatment effect (ATE) is given by

\[ \theta = E (Y_1 - Y_0) = p_1 - p_0. \]

• Suppose without lack of generality that \( q_0 \leq q_1 \). Then we can distinguish three subpopulations depending on an individual’s value of \( V \):

  – Always-takers: Units with \( V \leq q_0 \). They have \( D_1 = 1 \) and \( D_0 = 1 \). Their mass is \( q_0 \).
  – Compliers: Units with \( q_0 < V \leq q_1 \). They have \( D_1 = 1 \) and \( D_0 = 0 \). Their mass is \( q_1 - q_0 \).
  – Never-takers: Units with \( V > q_1 \). They have \( D_1 = 0 \) and \( D_0 = 0 \). Their mass is \( 1 - q_1 \).

• Note that membership of these subpopulations is unobservable, but we observe their mass.

• The local ATE (or LATE) is the average treatment effect for the subpopulation of compliers:

\[ \theta_{LATE} = E (Y_1 - Y_0 \mid q_0 < V \leq q_1). \]

• We have

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

and similarly

\[ E (Y_0 \mid q_0 < V \leq q_1) = \Pr (U_0 \leq p_0 \mid q_0 < V \leq q_1) = \frac{\Pr (p_0, q_1) - \Pr (p_0, q_0)}{q_1 - q_0}. \]

• Thus, the LATE satisfies a difference in differences expression of the form

\[ \theta_{LATE} = \frac{[C_1 (p_1, q_1) - C_1 (p_1, q_0)] - [C_0 (p_0, q_1) - C_0 (p_0, q_0)]}{q_1 - q_0} \]
3 Links with instrumental variable parameters

- Under monotonicity between \( D \) and \( Z \) (which the model assumes), \( \theta_{LATE} \) coincides with the Wald parameter (Imbens and Angrist, 1994):
  \[
  \theta_{LATE} = \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{E(D \mid Z = 1) - E(D \mid Z = 0)}
  \]

- To verify this result in our example, simply note that
  \[
  E(Y \mid Z = 1) = \Pr(Y = 1, D = 1 \mid Z = 1) + \Pr(Y = 1, D = 0 \mid Z = 1)
  = \Pr(U_1 \leq p_1, V \leq q_1) + \Pr(U_0 \leq p_0, V > q_1)
  = \Pr(U_1 \leq p_1, V \leq q_1) + \Pr(U_0 \leq p_0) - \Pr(U_0 \leq p_0, V \leq q_1)
  = C_1(p_1, q_1) + p_0 - C_0(p_0, q_1)
  \]

  \[
  E(Y \mid Z = 0) = \Pr(Y = 1, D = 1 \mid Z = 0) + \Pr(Y = 1, D = 0 \mid Z = 0)
  = \Pr(U_1 \leq p_1, V \leq q_0) + \Pr(U_0 \leq p_0, V > q_0)
  = \Pr(U_1 \leq p_1, V \leq q_0) + \Pr(U_0 \leq p_0) - \Pr(U_0 \leq p_0, V \leq q_0)
  = C_1(p_1, q_0) + p_0 - C_0(p_0, q_0)
  \]

  \[
  E(D \mid Z = 1) = E(D_1) = q_1, \quad E(D \mid Z = 0) = E(D_0) = q_0
  \]

- Moreover, \( E(Y_1 \mid q_0 < V \leq q_1) \) and \( E(Y_0 \mid q_0 < V \leq q_1) \) can also be calculated from Wald parameters (Abadie, 2002):
  \[
  E(Y_1 \mid q_0 < V \leq q_1) = \frac{E(Y_1 \mid Z = 1) - E(Y_1 \mid Z = 0)}{E(D_1 \mid Z = 1) - E(D_1 \mid Z = 0)}
  \]

  \[
  E(Y_0 \mid q_0 < V \leq q_1) = \frac{E(Y_0 \mid 1 - D \mid Z = 1) - E(Y_0 \mid 1 - D \mid Z = 0)}{E(1 - D_1 \mid Z = 1) - E(1 - D_1 \mid Z = 0)}
  \]

- To verify these results in our example note that
  \[
  E(YD \mid Z = 1) = \Pr(Y = 1, D = 1 \mid Z = 1) = \Pr(U_1 \leq p_1, V \leq q_1) = C_1(p_1, q_1)
  \]
  \[
  E(YD \mid Z = 0) = \Pr(Y = 1, D = 1 \mid Z = 0) = \Pr(U_1 \leq p_1, V \leq q_0) = C_1(p_1, q_0)
  \]

  and

  \[
  E(1 - D \mid Z = 1) - E(1 - D \mid Z = 0) = q_0 - q_1,
  \]

  \[
  E[Y(1 - D) \mid Z = 1] = \Pr(Y = 1, D = 0 \mid Z = 1)
  = \Pr(U_0 \leq p_0) - \Pr(U_0 \leq p_0, V \leq q_1) = p_0 - C_0(p_0, q_1)
  \]

  \[
  E[Y(1 - D) \mid Z = 0] = \Pr(Y = 1, D = 0 \mid Z = 0)
  = \Pr(U_0 \leq p_0) - \Pr(U_0 \leq p_0, V \leq q_0) = p_0 - C_0(p_0, q_0)
  \]

  \[
  3
  \]
4 Identification and estimation

- In conclusion, the mapping between reduced form and structural parameters is as follows. We observe $q_0$, $q_1$ and:

\[ E( Y D \mid Z = 1 ) = C_1 ( p_1, q_1 ) \tag{1} \]
\[ E( Y D \mid Z = 0 ) = C_1 ( p_1, q_0 ) \tag{2} \]

\[ E[ Y (1 - D) \mid Z = 1 ] = p_0 - C_0 ( p_0, q_1 ) \tag{3} \]
\[ E[ Y (1 - D) \mid Z = 0 ] = p_0 - C_0 ( p_0, q_0 ) \tag{4} \]

- Moreover, we know that:

\[ E( Y_1 \mid q_0 < V \leq q_1 ) = \frac{C_1 ( p_1, q_1 ) - C_1 ( p_1, q_0 )}{q_1 - q_0} \]
\[ E( Y_0 \mid q_0 < V \leq q_1 ) = \frac{C_0 ( p_0, q_1 ) - C_0 ( p_0, q_0 )}{q_1 - q_0} \]

- If $C_1 ( u, v )$ and $C_0 ( u, v )$ are Gaussian copulas with correlation coefficients $r_1$ and $r_0$, it turns out that $p_1$ and $r_1$ are just identified from (1)-(2), whereas $p_0$ and $r_0$ are just identified from (3)-(4). Thus, the switching regression probit model is just identified. So normality is not testable in this model, it is just an identifying assumption. However, if $U_1 \equiv U_0$ then the bivariate probit model 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.

- The representation (1)-(4) suggests a three-step estimation procedure:

  - Step 1: Estimate the “first-stage equation” to obtain $\hat{q}_0$ and $\hat{q}_1$ (or a more general propensity score if $Z$ has a larger support).
  - Step 2: Run a non-linear regression of $YD$ on $\hat{q}_Z$ using the copula model to estimate $p_1$ and any copula parameter in $C_1 ( u, v )$.
  - Step 3: Run a non-linear regression of $Y (1 - D)$ on $\hat{q}_Z$ using the copula model to estimate $p_0$ and any copula parameter in $C_0 ( u, v )$. 

4