Topics in Econometrics: Identification

Presented by Andrew Adrian Yu Pua

Universität Passau

April 7, 2016

1. Sources of identification failure
2. Links among instrumental variables, panel data, and simultaneous equations
3. Even in parametric settings, partial identification can be an outcome.
4. Identification can be possible through functional form restrictions and nonlinearity.
5. Binary choice models are tough to crack.
6. More examples today: Which assumptions are useful? Talking about which assumptions drive identification may be useful in improving conversations about applied work.

1. In semiparametric and nonparametric settings, normalizations are typically required.
2. Consider a binary choice model $\mathbb{E}(y|x) = \Pr(y = 1|x) = \Pr(\alpha + x\beta + \epsilon > 0|x)$.
3. Note that

$$\begin{aligned} \Pr(y = 1|x) &= Pr(\epsilon > -\alpha - x\beta|x) \\ &= \Pr\left(\frac{\epsilon - \mathbb{E}(\epsilon)}{var(\epsilon)} > \frac{-\alpha - x\beta - \mathbb{E}(\epsilon)}{var(\epsilon)}\middle| x\right) \\ &= \Pr(\epsilon^* > -\alpha^* - x\beta^*|x) \end{aligned}$$

4. In parametric binary choice models, typically impose $\mathbb{E}(\epsilon) = 0$ and $var(\epsilon) = 1$.
5. In semiparametric models, impose $\alpha = 0$ and $||\beta|| = 1$. Or impose zero mean and one of the coefficients in β to be equal to 1.

1. Let $\mathbb{E}(y|x,v) = \Pr(y=1|x,v) = \Pr(x\beta + \delta v + \epsilon > 0|x,v)$.
2. Make additional assumptions for semiparametric identification:

SBR-1 For $\alpha \in (0,1)$, $\Pr(\epsilon \leq 0|x,v) = \alpha$.
SBR-2 $\epsilon \perp (v_0, v_1)|(x,v)$
SBR-3 $\delta > 0$

3. Set $\delta = 1$ as normalization.
4. Let Assumption I and Assumptions SBR-1 to SBR-3 hold. Let $b \in \mathbb{R}^K$. Define

$$T(b) = \{(x, v_0, v_1) : \; xb + v_1 \leq 0 < x\beta + v_0\} \cup \{(x, v_0, v_1) : \; x\beta + v_1 \leq 0 < xb + v_0\}.$$

Then β is identified relative to b iff $\Pr(T(b)) > 0$.

5. (Special case) Suppose $v_0 = v_1$. Let Assumption I and Assumptions SBR-1 to SBR-3 hold. Let $b \in \mathbb{R}^K$. Define

$$T(b) = \{(x, v) : \; xb + v \leq 0 < x\beta + v\} \\ \cup \{(x, v) : \; x\beta + v \leq 0 < xb + v\}.$$

   Then β is identified relative to b iff $\Pr(T(b)) > 0$.

6. Let $B^* = \{b \in \mathbb{R}^K : \; \Pr(T(b)) = 0\}$. Assume that
   - There exists no proper linear subspace of $\mathbb{R}^K$ having probability one under $F(x)$.
   - $\Pr(a_0 \leq v_0 \leq v_1 \leq a_1 | x) > 0$ for all $(a_0, a_1) \in \mathbb{R}^2$ such that $a_0 < a_1$, a.e. x.

   Then $B^* = \{\beta\}$.

7. (Special case) Change to "$\Pr(a_0 \leq v \leq a_1 | x) > 0$ for all $(a_0, a_1) \in \mathbb{R}^2$ such that $a_0 < a_1$, a.e. x."

1. Quantile independence is needed. Mean independence is not enough to guarantee identification. But, ...
2. Consider the same binary choice model where $y = 1(\alpha + v + \epsilon \geq 0)$. As a consequence,

$$E(y|v) = \Pr(y = 1|v) = \Pr(\alpha + v + \epsilon > 0|v)$$

3. Assume $\epsilon \perp v$ and $\mathbb{E}(\epsilon) = 0$. Then,

$$E(y|v) = \Pr(y = 1|v) = \Pr(\alpha + v + \epsilon > 0|v) = F_{-\alpha-\epsilon}(v)$$

4. As a result,

$$-\int v \frac{\partial E(y|v)}{\partial v}\, dv = -\int v \frac{\partial F_{-\alpha-\epsilon}(v)}{\partial v}\, dv = -\mathbb{E}(-\alpha - \epsilon) = \alpha$$

We now show that

$$\mathbb{E}\left(\frac{y-1(v>0)}{f_v(v)}\right)=\alpha.$$

Note that

$$\begin{aligned}
\mathbb{E}\left(\frac{y-1(v>0)}{f_v(v)}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{y-1(v>0)}{f_v(v)}\middle| v\right)\right) \\
&= \mathbb{E}\left(\frac{\mathbb{E}(y|v)-1(v>0)}{f_v(v)}\right) \\
&= \int [\mathbb{E}(y|v)-1(v>0)]\ dv \\
&= v[\mathbb{E}(y|v)-1(v>0)]-\int v\frac{\partial E(y|v)}{\partial v}\ dv \\
&= v[F_{-\alpha-\epsilon}(v)-1(v>0)]+\alpha \\
&= \alpha
\end{aligned}$$

1. Consider the model

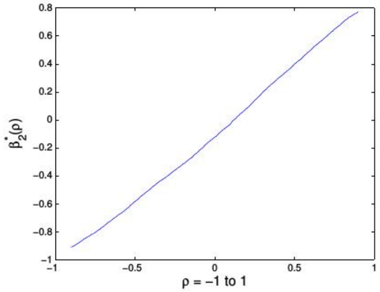
$$\begin{aligned} Y_1 &= 1(\beta_1 + u_1 > 0) \\ Y_2 &= 1(\beta_2 + \delta Y_1 + u_2 > 0) \end{aligned}$$

2. (u_1, u_2) is iid bivariate normal with mean $(0,0)$, $var(u_1) = var(u_2) = 1$, and $cov(u_1, u_2) = \rho \in (-1, 1)$
3. (Y_1, Y_2) is observable.
4. What can be identified in this case?
5. What if we have

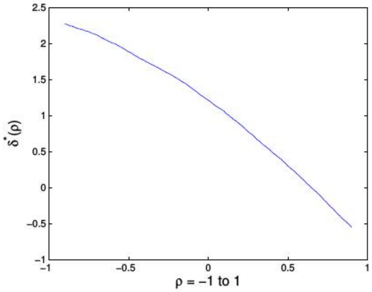
$$\begin{aligned} Y_1 &= 1(\beta_{11} + \beta_{12}x + u_1 > 0) \\ Y_2 &= 1(\beta_{21} + \beta_{22}x + \delta Y_1 + u_2 > 0) \end{aligned}$$

where $(u_1, u_2) \perp x$.

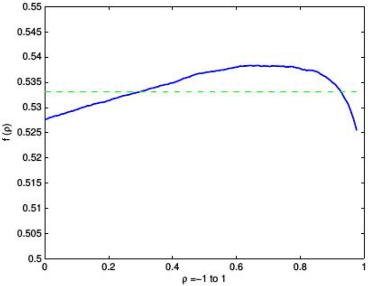
Bivariate probit II



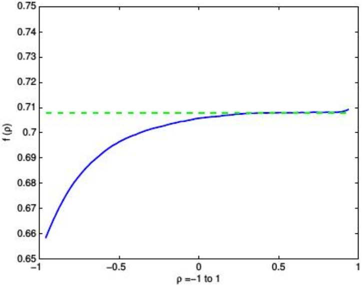
(a) (ρ, β_2) region



(b) (ρ, δ) region



(a) $(\rho, f(\rho))$ region, $\beta_{21} = -0.4$, $\rho_0 = -0.3$



(b) (ρ, δ) region, $\beta_{21} = 0.4$, $\rho_0 = 0.5$

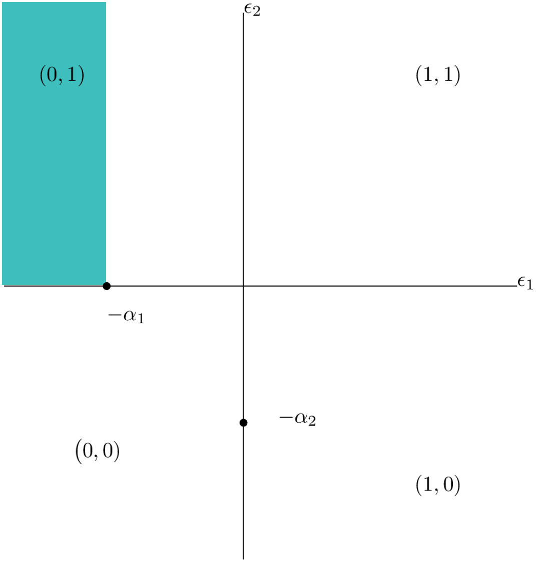
- Consider the following model where

$$y_1^* = \alpha_1 y_2 + \epsilon_1, y_1 = \mathbf{1}(y_1^* > 0)$$
$$y_2^* = \alpha_2 y_1 + \epsilon_2, y_2 = \mathbf{1}(y_2^* > 0)$$

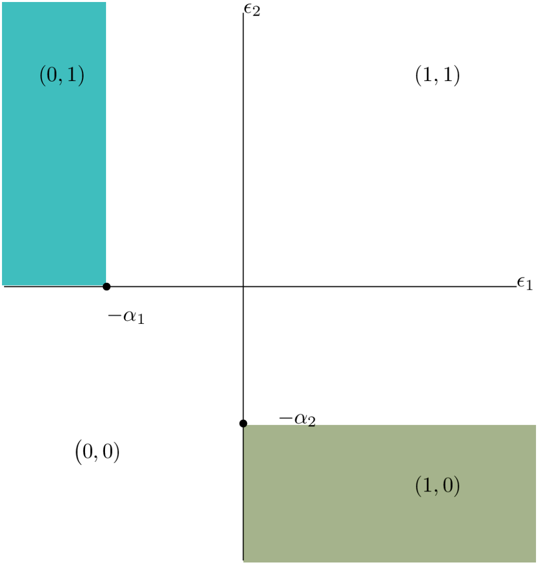
- Note that there are four possible observed outcomes.

$$(y_1, y_2) = \begin{cases} (1,1) & \text{if } \epsilon_1 > -\alpha_1, \epsilon_2 > -\alpha_2 \\ (1,0) & \text{if } \epsilon_1 > 0, \epsilon_2 \leq -\alpha_2 \\ (0,1) & \text{if } \epsilon_1 \leq -\alpha_1, \epsilon_2 > 0 \\ (0,0) & \text{if } \epsilon_1 \leq 0, \epsilon_2 \leq 0 \end{cases}$$

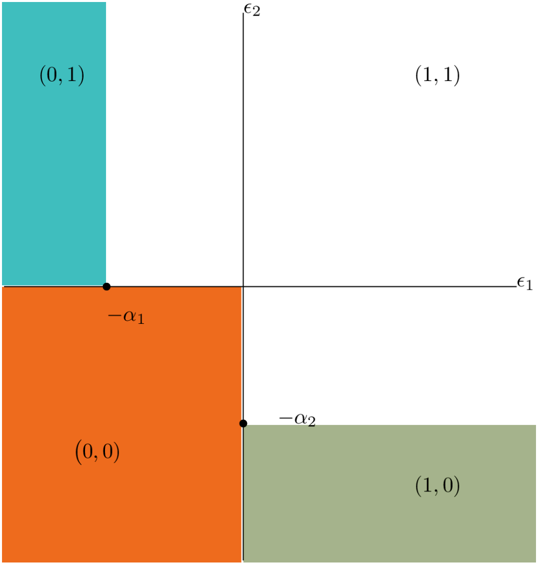
Visualizing $\alpha_1 > 0$ and $\alpha_2 > 0$



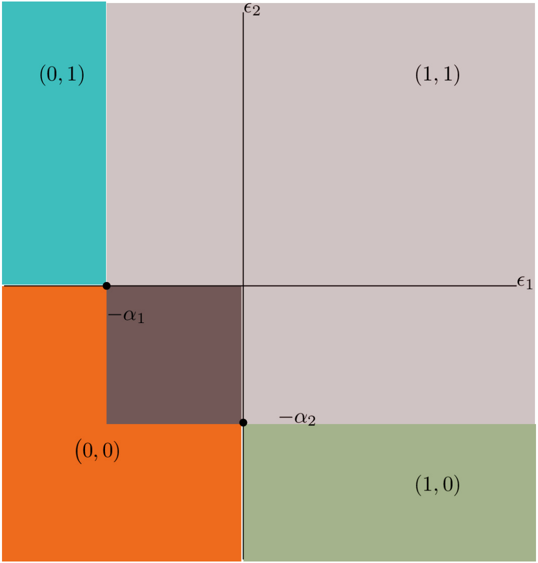
Visualizing $\alpha_1 > 0$ and $\alpha_2 > 0$

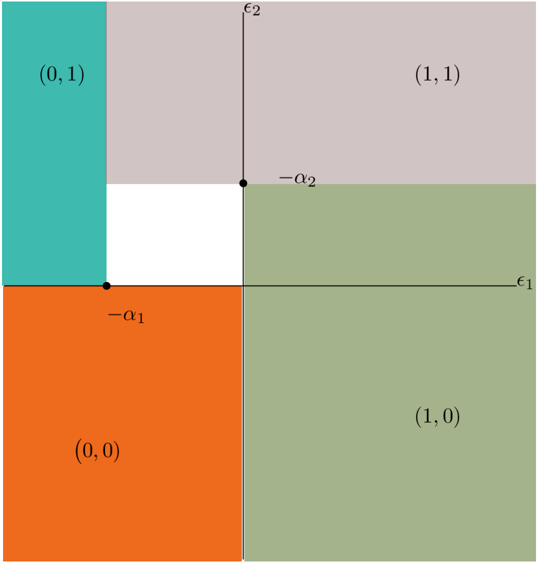


Visualizing $\alpha_1 > 0$ and $\alpha_2 > 0$

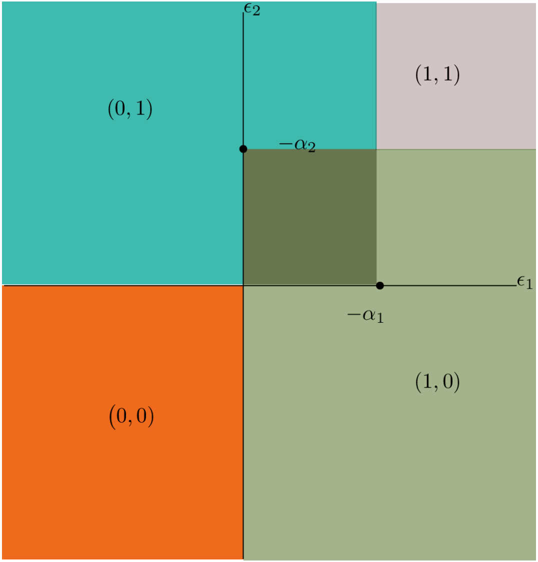


Visualizing $\alpha_1 > 0$ and $\alpha_2 > 0$




ϵ_2
$(0, 1)$
$(1, 1)$
$-\alpha_2$
ϵ_1
$-\alpha_1$
$(1, 0)$
$(0, 0)$

Visualizing $\alpha_1 < 0$ and $\alpha_2 < 0$



Tamer's (2003) solution

Consider the model where

$$y_1^* = x_1\beta_1 + y_2\alpha_1 + \epsilon_1$$
$$y_2^* = x_2\beta_2 + y_1\alpha_2 + \epsilon_2$$

and $y_1 = 1(y_1^* \geq 0)$ and $y_2 = 1(y_2^* \geq 0)$.

Assume that we have random sampling, $(\epsilon_1, \epsilon_2) \perp (x_1, x_2)$, (ϵ_1, ϵ_2) has a known continuous distribution with mean 0 and covariance matrix Ω, and that $\alpha_1 < 0, \alpha_2 < 0$.

If, in addition, we assume that for $i = 1$ or $i = 2$, there exists a regressor x_{ik} with $\beta_{ik} \neq 0$ such that $x_{ik} \notin x_{3-i}$ and such that the distribution of $x_{ik}|(x_{1k}, \ldots, x_{1,k-1}, x_{1,k+1}, \ldots, x_{1K})$ has an everywhere positive Lebesgue density. Then $(\beta_1, \beta_2, \alpha_1, \alpha_2)$ are identified if x_1 and x_2 have full rank.

Binary choice models with binary endogenous regressors I

1. Suppose we have a binary choice model (which can be rewritten as a nonseparable model):

$$Y = h(X, U) = \begin{cases} 0 & \text{if} \quad U \leq p(X) \\ 1 & \text{if} \quad p(X) \leq 1 \end{cases}$$

   where $U \sim U(0,1)$, $X \in \{x_1, x_2\}$ is endogenous, and Z is available as an instrumental variable, i.e. $U \perp Z$.

2. We want to identify the function $p(X)$. Let $\gamma_1 = p(X_1)$ and $\gamma_2 = p(X_2)$.
3. If $X \perp U$, then we have point identification.
4. Can an IV approach identify γ_1, γ_2?

Binary choice models with binary endogenous regressors II

5. Note that

$$
\begin{aligned}
\gamma_1 &= \Pr(U \leq \gamma_1) \\
&= \Pr(U \leq \gamma_1 | Z) \\
&= \Pr(U \leq \gamma_1 | X = x_1, Z) \Pr(X = x_1 | Z) \\
&\quad + \Pr(U \leq \gamma_1 | X = x_2, Z) \Pr(X = x_2 | Z) \\
&= \Pr(Y = 0 | X = x_1, Z) \Pr(X = x_1 | Z) \\
&\quad + \underbrace{\Pr(U \leq \gamma_1 | X = x_2, Z)}_{\in[0,1]} \Pr(X = x_2 | Z)
\end{aligned}
$$

6. Similarly,

$$
\begin{aligned}
\gamma_2 &= \underbrace{\Pr(U \leq \gamma_2 | X = x_1, Z)}_{\in[0,1]} \Pr(X = x_1 | Z) \\
&\quad + \Pr(Y = 0 | X = x_2, Z) \Pr(X = x_2 | Z)
\end{aligned}
$$

Binary choice models with binary endogenous regressors III

7. Bounds for both γ_1 and γ_2 are now available.
8. These bounds depend on Z. More importantly, they intersect each other.
9. Everything so far is completely nonparametric. What if we include a shape restriction? Suppose $\gamma_1 \leq \gamma_2$. Bounds can be tightened.
10. Note

$$\begin{aligned}
\gamma_1 &= \Pr(Y=0|X=x_1,Z)\Pr(X=x_1|Z) \\
&\quad + \Pr(U \leq \gamma_1|X=x_2,Z)\Pr(X=x_2|Z) \\
&\leq \Pr(Y=0|X=x_1,Z)\Pr(X=x_1|Z) \\
&\quad + \Pr(U \leq \gamma_2|X=x_2,Z)\Pr(X=x_2|Z) \\
&= \Pr(Y=0|X=x_1,Z)\Pr(X=x_1|Z) \\
&\quad + \Pr(Y=0|X=x_2,Z)\Pr(X=x_2|Z)
\end{aligned}$$

Binary choice models with binary endogenous regressors IV

11. Similarly,

$$\begin{aligned}\gamma_2 &= \Pr(U \leq \gamma_2 | X = x_1, Z) \Pr(X = x_1 | Z) \\ &\quad + \Pr(Y = 0 | X = x_2, Z) \Pr(X = x_2 | Z) \\ &\geq \Pr(Y = 0 | X = x_1, Z) \Pr(X = x_1 | Z) \\ &\quad + \Pr(Y = 0 | X = x_2, Z) \Pr(X = x_2 | Z)\end{aligned}$$

12. What if we impose a parametric restriction? Suppose

$$p(X) = \frac{1}{1 + \exp(\pi_0 + \pi_1 X)}$$

13. Note that

$$\pi_0 + \pi_1 x_1 = \log\left(\frac{1 - \gamma_1}{\gamma_1}\right)$$

$$\pi_0 + \pi_1 x_2 = \log\left(\frac{1 - \gamma_2}{\gamma_2}\right)$$

Final remarks

1. Hope you learned a lot from this short course.
2. I decided to emphasize examples and arguments rather than dig deep into one particular topic. I also wanted to emphasize commonalities in the examples.
3. There are many things still do study – inference, computation, and big data concerns.
4. The need for privacy will definitely change the nature of data collection assumptions.
5. The need for generating some "certainty" in the effects of policies will also change the nature of data generation assumptions.
6. Perhaps we should also search for the minimal set of assumptions for identification.
7. Finally, links between model misspecification and partial identification may have to be explored.