Lecture 7 – Identification in structural models

Economics 8379
George Washington University

Instructor: Prof. Ben Williams

Identification in parametric models

Del Boca et al. (2014)

Semiparametric and nonparametric identification

Partial identification

Identification

- Let $f_{Y|X}(Y_i \mid X_i; \theta)$ denote the conditional density of Y_i given X_i, which depends on a parameter vector θ, which must be in a parameter space Θ.
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.

Identification

- Let $f_{Y|X}(Y_i \mid X_i; \theta)$ denote the conditional density of Y_i given X_i, which depends on a parameter vector θ, which must be in a parameter space Θ.
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.
- The model is identified if this holds for any possible value of $\theta_0 \in \Theta$.
- The model is *locally identified* if this is true only for values of θ in a neighborhood around θ_0.

Identification

- If $\theta = (\theta_1, \theta_2)$ then it may be that θ_1 is identified but not θ_2.

Identification

- If $\theta = (\theta_1, \theta_2)$ then it may be that θ_1 is identified but not θ_2.
    - If $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ then $\theta_1 = \theta_{10}$.

Identification

- If $\theta = (\theta_1, \theta_2)$ then it may be that θ_1 is identified but not θ_2.
  - If $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ then $\theta_1 = \theta_{10}$.
- The model is partially identified if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta \in \Theta_0 \subseteq \Theta$
  - Θ_0 is called the “identified set”.

Identification

- It is *often* the case that we have to restrict Θ in order for the model to be identified.
  - Sometimes these restrictions have testable implications.
  - Sometimes they don't.
- If θ_0 is "close" to a point outside Θ where the model is not identified, inference may be effected.
  - This is called "weak identification"
- We will see an example of this shortly.

Maximum likelihood

- The MLE of θ is $\arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log \left(f_{Y|X}(Y_i \mid X_i; \theta) \right)$.
  - The estimator is consistent if $\mathcal{L}_n(\theta) \to_p \mathcal{L}(\theta)$ and $\mathcal{L}(\theta) = \mathcal{L}(\theta_0)$ implies that $\theta = \theta_0$.

Maximum likelihood

- The MLE of θ is $\arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log\left(f_{Y|X}(Y_i \mid X_i;\theta)\right)$.
  - The estimator is consistent if $\mathcal{L}_n(\theta) \to_p \mathcal{L}(\theta)$ and $\mathcal{L}(\theta) = \mathcal{L}(\theta_0)$ implies that $\theta = \theta_0$.
  - If the model is identified, the likelihood is correctly specified, and $\mathcal{L}(\theta) = E\left(\log(f_{Y|X}(Y_i \mid X_i;\theta))\right)$ then this condition holds.

Maximum likelihood

- A failure of local identification can often be detected by observing that the likelihood function is flat in some region.
- (Global) identification failures are harder to detect empirically.
- In complex models that require numerical integration or simulation, even local identification failures may be overlooked, particularly when θ is high-dimensional.
- Therefore, generally, identification should be proved theoretically.

MLE example

- Suppose that

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$
$$x_{1i} = \gamma_0 + \gamma_1 x_{2i} + v_i$$

  and $(u_i, v_i) \mid x_{2i} \sim N(0, \Sigma)$.
- Consider the likelihood with $Y_i = (y_i, x_{1i})$ and $X_i = x_{2i}$.
  - Let $\theta = (\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1, \Sigma)$.
  - The distribution is a bivariate normal with mean $\mu(x_{2i}; \theta) = (\beta_0 + \beta_1\gamma_0 + \beta_1\gamma_1 x_{2i} + \beta_2 x_{2i}, \gamma_0 + \gamma_1 x_{2i})$ and variance

$$\Sigma = \begin{pmatrix} \sigma_u^2 + \beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv} & \beta_1\sigma_v^2 + \sigma_{uv} \\ \beta_1\sigma_v^2 + \sigma_{uv} & \sigma_v^2 \end{pmatrix}$$

MLE example

- Identification failure:
  - For any $s \in [0, 1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1 - s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.

MLE example

- Identification failure:
  - For any $s \in [0,1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1-s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.
  - Then if we define β_0^s so that $\beta_0^s + \beta_1^s\gamma_0 = \beta_0 + \beta_1\gamma_0$

MLE example

- Identification failure:
  - For any $s \in [0, 1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1 - s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.
  - Then if we define β_0^s so that $\beta_0^s + \beta_1^s\gamma_0 = \beta_0 + \beta_1\gamma_0$
  - and σ_{uv}^s so that $\beta_1^s\sigma_v^2 + \sigma_{uv}^s = \beta_1\sigma_v^2 + \sigma_{uv}$

MLE example

- Identification failure:
  - For any $s \in [0,1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1-s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.
  - Then if we define β_0^s so that $\beta_0^s + \beta_1^s\gamma_0 = \beta_0 + \beta_1\gamma_0$
  - and σ_{uv}^s so that $\beta_1^s\sigma_v^2 + \sigma_{uv}^s = \beta_1\sigma_v^2 + \sigma_{uv}$
  - and σ_u^s so that $(\sigma_u^s)^2 + (\beta_1^s)^2\sigma_v^2 + 2\beta_1^s\sigma_{uv}^s = \sigma_u^2 + \beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}$

MLE example

- Identification failure:
  - For any $s \in [0,1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1-s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.
  - Then if we define β_0^s so that $\beta_0^s + \beta_1^s\gamma_0 = \beta_0 + \beta_1\gamma_0$
  - and σ_{uv}^s so that $\beta_1^s\sigma_v^2 + \sigma_{uv}^s = \beta_1\sigma_v^2 + \sigma_{uv}$
  - and σ_u^s so that $(\sigma_u^s)^2 + (\beta_1^s)^2\sigma_v^2 + 2\beta_1^s\sigma_{uv}^s = \sigma_u^2 + \beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}$
  - then $\mathcal{L}(\theta^(s)) = \mathcal{L}(\theta)$.

MLE example

- Identification failure:
  - For any $s \in [0,1]$ define $\beta_1^s = s(\beta_1 + \beta_2/\gamma_1)$ and $\beta_2^s = (1-s)(\beta_1\gamma_1 + \beta_2)$.
  - Then notice that $\beta_1^s\gamma_1 + \beta_2^s = \beta_1\gamma_1 + \beta_2$ for every s.
  - Then if we define β_0^s so that $\beta_0^s + \beta_1^s\gamma_0 = \beta_0 + \beta_1\gamma_0$
  - and σ_{uv}^s so that $\beta_1^s\sigma_v^2 + \sigma_{uv}^s = \beta_1\sigma_v^2 + \sigma_{uv}$
  - and σ_u^s so that $(\sigma_u^s)^2 + (\beta_1^s)^2\sigma_v^2 + 2\beta_1^s\sigma_{uv}^s = \sigma_u^2 + \beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}$
  - then $\mathcal{L}(\theta^(s)) = \mathcal{L}(\theta)$.
- Solutions:
  - assume that $\sigma_{uv} = 0$ (i.e., X_{1i} is exogenous)
  - *or* assume that $\beta_2 = 0$ (i.e., an exclusion restriction)

MLE example

- Consider the second solution that $\beta_2 = 0$.
  - Under this restriction, the model is identified if $\gamma_1 \neq 0$.
  - Even after imposing $\beta_2 = 0$, we have to restrict Θ further to get identification.
    - This is an example of a testable restriction.
  - But what happens if β_2 is close to 0?

MLE example

- Consider the second solution that $\beta_2 = 0$.
  - Under this restriction, the model is identified if $\gamma_1 \neq 0$.
  - Even after imposing $\beta_2 = 0$, we have to restrict Θ further to get identification.
    - This is an example of a testable restriction.
  - But what happens if β_2 is close to 0?
    - *weak identification*

GMM

- Consider GMM estimation of θ based on the moment conditions, $E(w_i m(y_i, x_i, \theta_0)) = 0$.
  - Then the model is identified at θ_0 if $E(w_i m(y_i, x_i, \theta)) = 0$ implies that $\theta = \theta_0$.
  - The model is identified if this holds regardless of the value of θ_0.
  - The model is *locally identified* if this is true only for values of θ in a neighborhood around θ_0.

GMM

- Consider GMM estimation of θ based on the moment conditions, $E(w_i m(y_i, x_i, \theta_0)) = 0$.
    - Then the model is identified at θ_0 if $E(w_i m(y_i, x_i, \theta)) = 0$ implies that $\theta = \theta_0$.
    - The model is identified if this holds regardless of the value of θ_0.
    - The model is *locally identified* if this is true only for values of θ in a neighborhood around θ_0.

    - n.b. It is possible that (a) $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$ but (b) there is some $\theta \neq \theta_0$ such that $E(w_i m(y_i, x_i, \theta)) = 0$.

Indirect inference

- Last class we also saw a condition for identification based on indirect inference (i.e., the method of simulated moments).
  - The binding function $\theta(\beta)$ must be one-to-one.
- This is often hard to verify.
- Del Boca, Flinn, Wiswall (2014, REStud) provides a good example.

Del Boca et al. (2014) model

- Parents choose time ($\tau_{1,t}$ and $\tau_{2,t}$) and goods (e_t) to invest in child, along with leisure and work hours ($l_{j,t}$ and $h_{j,t}$, $j = 1, 2$) .
  - Child quality: $k_{t+1} = R_t \tau_{1,t}^{\delta_{1,t}} \tau_{2,t}^{\delta_{2,t}} e_t^{\delta_{3,t}} k_t^{\delta_{4,t}}$.
  - Period utility: $u = \alpha_1 \ln l_{1t} + \alpha_2 \ln l_{2t} + \alpha_3 \ln c_t + \alpha_4 \ln k_t$
  - Period budget constraint: $c_t + e_t = w_{1t} h_{1t} + w_{2t} h_{2t} + I_t$
  - Time constraint: $T = l_{j,t} + h_{j,t} + \tau_{j,t}$ for each $j = 1, 2$

Del Boca et al. (2014) model

- Finite horizon optimization.
  - There are M time periods (ages) and parents maximize expected discounted sum of utility over the M periods.
  - This requires specification of a terminal value.
  - The model can then be solved by backwards induction.

Del Boca et al. (2014) model

- Solution.
  - $\tau_{j,t} = (T - h_{j,t})\frac{\phi_{j,t}}{\alpha_j + \phi_{j,t}}$ and $e_t = (w_{1t}h_{1t} + w_{2t}h_{2t} + I_t)\frac{\phi_{3,t}}{\alpha_3 + \phi_{3,t}}$ where $\phi_{j,t} = \beta\delta_{j,t}\eta_{t+1}$
  - The η_t are solved recursively as a function of α_4 and $\delta_{4,t}$.
  - Labor supply also has a convenient closed-form solution, though corner solutions needed to be accounted for.

Del Boca et al. (2014) econometric specifications

- Children are observed at different ages so specify

$$\delta_{j,t} = \exp(\gamma_{j,0} + \gamma_{j,1} t)$$

- The TFP sequence includes a stochastic component,

$$R_t = \exp(\gamma_{0,0} + \gamma_{0,1} t + \bar{\omega}_t)$$

- Individual hetereogeneity in α: distribution $G(\alpha)$ that enforces $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$
- Joint wage equations for spouses (education, age, age squared, year of birth, correlated errors).
- Censored process for non-labor income: $I_t = \max\{0, I_t^*\}$.

Del Boca et al. (2014) identification

- Suppose we observe each child i for two periods starting with period t_i.
- We could estimate: $\ln k_{i,t_i+1} = \gamma_{0,0} + \gamma_{0,1} t_i + \delta_{1,t_i} \ln \tau_{i,1,t_i} + \delta_{2,t_i} \ln \tau_{i,2,t_i} + \delta_{3,t_i} \ln e_{i,t_i} + \delta_{4,t_i} \ln k_{i,t_i} + \eta_{i,t_i}$

Del Boca et al. (2014) identification

- The wage equations are harder to estimate because of censoring. We only observe wage $w_{j,t}$ if $h_{j,t} > 0$.

Del Boca et al. (2014) identification

- The wage equations are harder to estimate because of censoring. We only observe wage $w_{j,t}$ if $h_{j,t} > 0$.
    - "Under our model specification, we can 'correct' our estimator of model parameters for the non-randomly missing data using the DGP structure from the model."

Del Boca et al. (2014) identification

- The wage equations are harder to estimate because of censoring. We only observe wage $w_{j,t}$ if $h_{j,t} > 0$.
  - "Under our model specification, we can 'correct' our estimator of model parameters for the non-randomly missing data using the DGP structure from the model."
  - " In this case, both the wage processes and the parameters characterizing preferences and production technologies must be simultaneously estimated."

Del Boca et al. (2014) identification

- Suppose we know β.
  - Given $\delta_{j,t}$'s we can estimate α *for each household* using the input demand equations.
  - $G(\alpha)$ is nonparametrically identified

Del Boca et al. (2014) identification

- Suppose we know β.
    - Given $\delta_{j,t}$'s we can estimate α *for each household* using the input demand equations.
    - $G(\alpha)$ is nonparametrically identified
- Suppose α is known.
    - Use the labor supply equations to estimate β.

Del Boca et al. (2014) identification

- This is *not* a rigorous identification argument, though for many it is “good enough”.
- Issues:
  - Is $G(\alpha)$ really identified?
  - The input demand and labor supply equations have to be linearly independent.
  - Labor supply censoring ...

Del Boca et al. (2014) identification

- Once we're convinced the model is identified, we also need to use moments (auxiliary models) that are sufficient for the binding function to be one-to-one.
- The identification argument should inform us in this regard.

Identification in parametric models

Del Boca et al. (2014)

Semiparametric and nonparametric identification

Partial identification

Semiparametric models

- A semiparametric model is one in which θ is composed of θ_1 and θ_2 where θ_1 is finite dimensional (i.e., a length K vector) and θ_2 is infinite-dimensional (i.e., a function, or a vector of functions).
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.

Semiparametric models

- A semiparametric model is one in which θ is composed of θ_1 and θ_2 where θ_1 is finite dimensional (i.e., a length K vector) and θ_2 is infinite-dimensional (i.e., a function, or a vector of functions).
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.

  example:
  - Y_i is binary and $Pr(Y_i = 1 \mid X_i) = F_\varepsilon(\beta' X_i)$.
  - Unless we assume that F_ε is known (for example, $F_\varepsilon = \Phi$), then $\theta_1 = \beta$ and $\theta_2 = F_\varepsilon$ and this is a semiparametric model.
  - This model is not identified without restrictions on F_ε and/or β.

Nonparametric models

- A nonparametric model is one in which θ is infinite-dimensional (i.e., a function, or a vector of functions).
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.

Nonparametric models

- A nonparametric model is one in which θ is infinite-dimensional (i.e., a function, or a vector of functions).
- The model is identified at θ_0 if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\theta = \theta_0$.

  example 1:

  - Y_i is binary and $Pr(Y_i = 1 \mid X_i) = m(X_i)$.
  - Here if we don't assume anything about the function $m(x)$ then this is a nonparametric model.
  - The function m is identified (at points x in the support of X_i).

  example 2:

  - Y_i is binary and $Pr(Y_i = 1 \mid X_i) = F_\varepsilon(m(X_i))$.
  - This is a random utility model where $m(x)$ and F_ε represent distinct parts of the underlying structural model. This is also a nonparametric model.
  - This model is not identified!

Identification of some features

- Sometimes some parameters in the model are identified and others are not.
  - In the bivariate normal model above, γ_0, γ_1 and σ_v are identified without any restrictions on Θ.
- Sometimes certain combinations of parameters are identified even when none are individually identified.
  - In the semiparametric RUM model above, if F_ε is differentiable with derivative f_ε then

$$\frac{\partial}{\partial x_k} Pr(Y_i = 1 \mid X_i = x) = \beta_k f_\varepsilon(\beta' x)$$

  - Therefore, β_k/β_l is identified as long as $\beta_l \neq 0$.
  - This result can be proved under weaker assumptions as well.

Identification of some features

- Both of these cases fit within the following definition.
  - A feature of the model, represented by a mapping $\psi(\theta)$, is identified if $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$ implies that $\psi(\theta) = \psi(\theta_0)$.

Partial identification

- The *identified set* for a parameter θ is

$$\Theta^I = \{\theta \in \Theta : f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)\}$$

- The identified set depends on θ_0 and Θ.
- The identified set for a feature of the model, $\psi(\theta)$ is

$$\psi(\Theta^I) = \{\psi(\theta) : \theta \in \Theta^I\}$$

Example 1

- Consider estimation of the *ATE* when the common support condition is not satisfied.
  - Let *CS* denote the common support.
  - Suppose $0 \leq Y_i \leq 1$ and that $Pr(X_i \in CS) = \pi$.
  - Note that

$$
\begin{aligned}
ATE &= \int \delta(x) f_X(x) dx \\
&= \int_{CS} \delta(x) f_X(x) dx + \int_{CS^c} \delta(x) dx
\end{aligned}
$$

Example 1

- First, $-(1-\pi) \leq \int_{CS^c} \delta(x)dx \leq 1-\pi$.
- Second, the ATE on the common support is $ATE_{CS} = \int_{CS} \delta(x)\frac{f_X(x)}{\pi}dx$.
- Therefore,

$$ATE_{CS}\pi - (1-\pi) \leq ATE \leq ATE_{CS}\pi + (1-\pi)$$

- If $\delta(x)$ is identified for $x \in CS$ then these two bounds are identified, meaning that the identified set for ATE is the interval $[ATE_{CS}\pi - (1-\pi), ATE_{CS}\pi + (1-\pi)]$.

Example 2

- Economic models with multiple equilibria only provide moment *in*equalities.
- Suppose that $E(w_i m(y_i, x_i, \theta_0)) \leq 0$.
- Then the identified set for θ is

$$\Theta^I = \{\theta \in \Theta : E(w_i m(y_i, x_i, \theta)) \leq 0\}$$

- See, e.g., Manski and Tamer (2002) and Ciliberto and Tamer (2009).

Bounds and sharp identified set

- As in example 1, suppose we can prove that $L \leq \psi(\theta_0) \leq U$.
  - It is not automatically the case that for every $z \in [L, U]$, there is a θ such that $z = \psi(\theta)$ and $f_{Y|X}(y \mid x; \theta) = f_{Y|X}(y \mid x; \theta_0)$.
  - When this distinction is important, sometimes $[L, U]$ is referred to as the identified set and $\psi(\Theta^I)$ as the *sharp* identified set.

Ciliberto and Tamer (2009)

- This is a game-theoretic model of entry into a market.
- Suppose there are two players with random best response functions,

$$y_{1m} = \mathbf{1}(\alpha_1 X_{1m} + \delta_2 y_{2m} + \varepsilon_{1m} \geq 0)$$
$$y_{2m} = \mathbf{1}(\alpha_2 X_{2m} + \delta_1 y_{1m} + \varepsilon_{2m} \geq 0)$$

for player i in market m.

Ciliberto and Tamer (2009)

- This is a game-theoretic model of entry into a market.
- Suppose there are two players with random best response functions,

$$y_{1m} = \mathbf{1}(\alpha_1 X_{1m} + \delta_2 y_{2m} + \varepsilon_{1m} \geq 0)$$
$$y_{2m} = \mathbf{1}(\alpha_2 X_{2m} + \delta_1 y_{1m} + \varepsilon_{2m} \geq 0)$$

  for player i in market m.
- For some values of $\delta_1, \delta_2, \alpha_1' X_{1im}$ and $\alpha_2' X_{2m}$, there is a unique Nash equilibrium.
- For other values of $\delta_1, \delta_2, \alpha_1' X_{1im}$ and $\alpha_2' X_{2m}$, there are multiple Nash equilibria.

Ciliberto and Tamer (2009)

- Player 1 enters and 0 does not is a Nash equilibrium if: $-\alpha_1 X_{1m} \leq \varepsilon_{1m}$ and $-\alpha_2 X_{2m} - \delta_1 \geq \varepsilon_{2m}$

Ciliberto and Tamer (2009)

- Player 1 enters and 0 does not is a Nash equilibrium if: $-\alpha_1 X_{1m} \leq \varepsilon_{1m}$ and $-\alpha_2 X_{2m} - \delta_1 \geq \varepsilon_{2m}$
- Player 0 enters and 1 does not is a Nash equilibrium if: $-\alpha_1 X_{1m} - \delta_2 \geq \varepsilon_{1m}$ and $-\alpha_2 X_{2m} \leq \varepsilon_{2m}$

Ciliberto and Tamer (2009)

- Player 1 enters and 0 does not is a Nash equilibrium if: $-\alpha_1 X_{1m} \leq \varepsilon_{1m}$ and $-\alpha_2 X_{2m} - \delta_1 \geq \varepsilon_{2m}$
- Player 0 enters and 1 does not is a Nash equilibrium if: $-\alpha_1 X_{1m} - \delta_2 \geq \varepsilon_{1m}$ and $-\alpha_2 X_{2m} \leq \varepsilon_{2m}$
- If $\delta_1, \delta_2 < 0$ then these regions overlap!!

Ciliberto and Tamer (2009)

- The probability that player 1 enters and 0 does not can be written as

$$
\begin{aligned}
&Pr((\varepsilon_1, \varepsilon_2) \in R_1(X, \theta)) \\
&+ \int \pi(\varepsilon_1, \varepsilon_2) \mathbf{1}((\varepsilon_1, \varepsilon_2) \in R_2(X, \theta)) dF_{\varepsilon_1, \varepsilon_2}
\end{aligned}
$$

for two distinct regions of $\mathbb{R}^2$, where $\pi(\varepsilon_1, \varepsilon_2)$ is an unspecified, unknown equilibrium selection rule.

Ciliberto and Tamer (2009)

- The probability that player 1 enters and 0 does not can be written as

$$Pr((\varepsilon_1, \varepsilon_2) \in R_1(X, \theta)) + \int \pi(\varepsilon_1, \varepsilon_2)\mathbf{1}((\varepsilon_1, \varepsilon_2) \in R_2(X, \theta))dF_{\varepsilon_1, \varepsilon_2}$$

  for two distinct regions of $\mathbb{R}^2$, where $\pi(\varepsilon_1, \varepsilon_2)$ is an unspecified, unknown equilibrium selection rule.
- Since $\pi(\varepsilon_1, \varepsilon_2)$ is a probability it has to be between 0 and 1 and therefore,

$$Pr((\varepsilon_1, \varepsilon_2) \in R_1(X, \theta)) \leq Pr((1, 0) \mid X)$$
$$\leq Pr((\varepsilon_1, \varepsilon_2) \in R_1(X, \theta)) + \int \mathbf{1}((\varepsilon_1, \varepsilon_2) \in R_2(X, \theta))dF_{\varepsilon_1, \varepsilon_2}$$

Moment inequality estimation

- One approach to estimating the identified set based on $E(w_i m(y_i, x_i, \theta_0)) \leq 0$:
  - Define
$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(w_i m(y_i, x_i, \theta) > 0) w_i m(y_i, x_i, \theta).$$
  - Then $\hat{\Theta}^I = \{\theta \in \Theta : Q_n(\theta) \leq \nu_n\}$ where ν_n is nuisance parameter that must be specified.
  - And a confidence set is given by $\{\theta \in \Theta : n\left(Q_n(\theta) - \min_t Q_n(t)\right) \leq c_\alpha\}$, where c_α is a critical value.