Topics in Econometrics: Identification

Presented by Andrew Adrian Yu Pua

Universität Passau

April 4, 2016

1. Conjures up memories of models involving simultaneous equations and instrumental variables; Demand-supply diagrams are the usual examples.
2. Early encounters already in the classical linear regression model
3. (Not so exciting, but ...) Let $Y_i = a + bx_i + \delta_i$ for $i = 1, \ldots, 100$. The x_i's are known and fixed at 2 for all i. Assume further that δ_i's are iid, $\mathbb{E}(\delta_i) = 0$, and $var(\delta_i) = \sigma^2$. Can we identify the parameters a, b, σ^2?
4. In slightly advanced courses, you may have encountered identification issues in binary choice models.
5. Bottom line: We want to characterize what can be learned about some parameter θ from observables.
6. Elements: Data collection assumptions and data generation assumptions

"Rough" definitions

Suppose P_θ is the probability distribution that governs an observable random variable X.

1. The function $f(\theta)$ is identifiable if

$$f(\theta_1) \neq f(\theta_2) \implies P_{\theta_1} \neq P_{\theta_2}.$$

   for every (θ_1, θ_2).

2. The function $f(\theta)$ is estimable if there exists a function g depending only on the observables such that

$$\mathbb{E}_\theta[g(X)] = f(\theta).$$

So, what is the relationship between identifiability and estimability? Let us work through Exercise Set C.

Why should we be concerned about identification?

1. One of the ingredients in a consistency proof (though obscured or reformulated most of the time)
2. Can be useful in the construction of estimators outside the likelihood framework
3. "Before any inferential procedure can be developed, one needs to assert that the unknown parameters are identifiable." (Basu, EOSS)
4. Pushed to the extreme: "If two parameter values imply the same distribution of the data, no observed data can distinguish between them, and there will be no point in attempting to estimate those parameters. Thus consideration of identification is a question that logically precedes estimation." (Schmidt, EOSS)

Newey and McFadden (1994) Theorem 2.1

Let $\widehat{Q}_n(\theta)$ be some objective function such that $\widehat{\theta}$ maximizes $\widehat{Q}_n(\theta)$ subject to $\theta \in \Theta$. If there is a function $Q_0(\theta)$ such that

1. $Q_0(\theta)$ is uniquely maximized at θ_0;
2. Θ is compact;
3. $Q_0(\theta)$ is continuous;
4. $\sup_{\theta \in \Theta} |\widehat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$,

then $\widehat{\theta} \xrightarrow{p} \theta_0$.

Likelihood and extremum based identification

1. (ML) Let $f(z|\theta)$ be some pdf. If
   - θ_0 is identified,
   - $\mathbb{E}[|\log f(z|\theta)|] < \infty$ for all θ,

   then
   $$Q_0(\theta) = \mathbb{E}[\log f(z|\theta)]$$
   has a unique maximum at θ_0.
2. (GMM) Let $g_0(\theta) = \mathbb{E}[g(z,\theta)]$ be some set of population moments. If
   - W is positive semidefinite,
   - $g_0(\theta_0) = 0$,
   - $Wg_0(\theta) \neq 0$ for $\theta \neq \theta_0$,

   then
   $$Q_0(\theta) = -g_0(\theta)' W g_0(\theta)$$
   has a unique maximum at θ_0.

1. (Multivariate regression) Let Y be a $k \times 1$ vector which is multivariate normal with mean $X\beta_0$ and nonsingular covariance matrix Σ_0 (general) or $\Sigma_0 = \sigma_0^2 I$.
2. (Probit) Let $z = (y, x')$ where $y \in \{0, 1\}$ and x is a $q \times 1$ vector of regressors. Consider the pdf

$$f(z|\theta) = [\Phi(x'\theta)]^y [1 - \Phi(x'\theta)]^{1-y},$$

   where Φ is the standard normal cdf.
3. (IV) Let $z = (x', y, Y')$, where x is a vector of instrumental variables, y is the dependent variable, and Y' is a vector of RHS endogenous variables.Consider the moment function

$$g(z, \theta) = x'(y - Y'\theta)$$

1. (Linear IV) Suppose $Y = \alpha + \beta X + U$ where $\mathbb{E}(U|Z) = c$. When can we identify β?
2. (A step to nonlinear IV regression) Suppose $Y = \alpha + \beta X + \gamma X^2 + U$ where $\mathbb{E}(U|Z) = c$. When can we identify β and γ?
3. (Nonparametric IV) Let $X \in \{x_1, \ldots, x_M\}$ and $Z = \{z_1, \ldots, z_K\}$. Suppose $Y = h(X) + U$ with $\mathbb{E}(U|Z) = 0$. Can we identify h?
4. (MA processes) Let $X_t = \mu + b_0\epsilon_t + b_1\epsilon_{t-1} + \ldots + b_p\epsilon_{t-p}$. Assume for the moment that $\epsilon_t \sim \text{iid } N(0, \sigma^2)$. We observe X_t but not the ϵ's. Can we identify the b's and μ? Take the case where $p = 1$.

5. (Simultaneous equations) Let

$$\begin{aligned} Y_1 &= \beta Y_2 + \gamma Z + \epsilon_1 \\ Y_2 &= \delta Y_1 + \epsilon_2 \end{aligned}$$

Suppose $\mathbb{E}(\epsilon_1 Z) = \mathbb{E}(\epsilon_2 Z) = 0$. Can we identify δ? How about β?

6. (Panel binary choice) Suppose (Y_1, Y_2, Y_3, α) is a random vector such that

$$\begin{aligned} \Pr(Y_1 = 1|\alpha) &= p_1(\alpha) \\ \Pr(Y_2 = 1|\alpha, Y_1) &= F(\alpha + \gamma Y_1) \\ \Pr(Y_3 = 1|\alpha, Y_1, Y_2) &= F(\alpha + \gamma Y_2) \end{aligned}$$

where p_1 and F are both unknown functions with range in $[0, 1]$. In addition, F is strictly increasing.

7. (Dynamic panel data) Arellano and Bond (1991) proposed linear moment conditions of the form

$$\mathbb{E}[y_{is}(\Delta y_{it} - \alpha \Delta y_{i,t-1})] = 0, \; s = 0, \ldots, t-2; t = 2, \ldots, T,$$

to identify α_0 in the dynamic panel data model $y_{it} = \alpha_0 y_{i,t-1} + c_i + v_{it}$ under the following assumptions:

- $\mathbb{E}(v_{it}) = 0, \; t = 1, \ldots, T$
- $\mathbb{E}(v_{it} y_{i0}) = 0, \; t = 1, \ldots, T$
- $\mathbb{E}(v_{it} c_i) = 0, \; t = 1, \ldots, T$
- $\mathbb{E}(v_{is} v_{it}) = 0, \; s = 1, \ldots, t-1; t = 1, \ldots, T$

When will these linear moment conditions fail to identify α_0?

8. (Dynamic panel data) What if one uses Ahn-Schmidt quadratic moment conditions?

$$\mathbb{E}[y_{iT} - y_{i,T-1}\alpha)(\Delta y_{it} - \alpha \Delta y_{i,t-1})] = 0, \; t = 2, \ldots, T$$

Pötscher and Prucha's (1997) Lemma 3.1

Let $R_n : \Omega \times B \to \mathbb{R}$ and $\overline{R}_n : B \to \mathbb{R}$ such that a.s.

$$\sup_B |R_n(\omega, \beta) - \overline{R}_n(\beta)| \to 0 \text{ as } n \to \infty.$$

Let $\overline{\beta}_n$ be an identifiably unique sequence of minimizers of $\overline{R}_n(\beta)$, then for any sequence $\widehat{\beta_n}$ such that eventually

$$R_n(\omega, \widehat{\beta_n}) = \inf_B R_n(\omega, \beta)$$

holds, we have $\rho_B(\widehat{\beta_n}, \overline{\beta}_n) \to 0$ a.s. as $n \to \infty$.

Remarks

1. If $\overline{R}_n$ does not depend on n and the parameter space B is compact, then identifiable uniqueness is equivalent to the existence of a unique minimizer of $\overline{R}$.
2. This only happens under correct specification.
3. Under misspecification, we might have a set of minimizers.
4. A reparametrization may be possible so that the lack of identification disappears.
5. See Section 4.6 of Pötscher and Prucha (1997) for more discussion of the identification conditions under misspecification.

Identifiability and consistency

1. Observe that if there exists a consistent estimator for a parameter θ, then the parameter is identifiable.
2. Gabrielsen (1978, JoE) gives a counterexample to show that the converse is not true.
3. Consider the model where $Y_i = \beta\rho^i + \epsilon_i$, $i = 1 \ldots, n$. Assume that ρ is known with $\rho < 1$, $\epsilon_i \sim \mathrm{iid}\ N(0, 1)$ and $\beta > 0$.
4. Note that β is identifiable from the first moment of Y_i.
5. Form the ML estimator for β and show that it is not consistent.

A different type of argument: Measurement error model

Consider the model:

$$\begin{aligned} Y_1 &= \eta_1 + \epsilon_1 \\ Y_2 &= \eta_2 + \epsilon_2 \end{aligned}$$

where

1. $\eta_2 = \alpha + \beta\eta_1$
2. $\eta_1 \perp (\epsilon_1, \epsilon_2)$
3. (ϵ_1, ϵ_2) is bivariate normal with mean $(0, 0)$ and covariance matrix Σ
4. Y_1 and Y_2 are observable, but η_1 and η_2 are unobservable

A classic result: β is identifiable iff (η_1, η_2) is not bivariate normal.