SPRING 2018

ECON 2140: ECONOMETRIC METHODS

PROFESSOR ELIE TAMER

Compiled by: Sagar Saxena, Roman Sigalov

Table of Contents

Chapter 1

Consistency and Normality of Extremum Estimators

1.1 Introduction

This chapter covers results on consistency and asymptotic normality of extremum estimators, and lays out the various assumptions that lead to these results. We also cover examples of *identification* under different estimators as presented in class.

1.2 Consistency

1.2.1 Extremum Estimators

Extremum estimators are defined as a sequence of estimators $\{\hat{\theta}_n\}_{n\geq 1}$ that *approximately* ***minimize*** a stochastic objective function $Q_n(\theta)$.

Assumption 1.2.1 (EE).

$$\hat{\theta}_n \in \Theta \textit{ and } Q_n(\hat{\theta}_n) \leq \inf_{\theta\in\Theta} Q_n(\theta) + o_p(1)$$

Here, θ is a d-dimensional vector in the parameter space $\Theta \subset \mathbb{R}^d$. $o_p(1)$ is a random variable that *converges in probability*[1] to 0 as $n \to \infty$. The idea of "approximate" minimization is captured by the ?

1.2.2 Uniform Convergence

This is a regularity condition on the objective function to ensure that this function behaves well in the limit.

[1]The sequence $\{X_n\}$ of random variables converges in probability to the random variable X if for all $\epsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0$

Assumption 1.2.2 (UCONV).

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \to^p 0 \textit{ for some function } Q(\theta)$$

1.2.2.A Necessary Conditions for Uniform Convergence

Assumption 1.2.3 (P-CONV[2]). *For all $\theta \in \Theta$, $Q_n(\theta) \to^p Q(\theta)$*

Lemma 1.1

Let the following hold: (i) $\{W_i\}_{i=1}^n$ i.i.d., (ii) $m(W_i, \theta)$ continuous in θ for all W_i, (iii) $\mathbb{E}[\sup_{\theta \in \Theta} |m(W_i, \theta)|] < \infty$, and (iv) Θ is compact. Then, uniform convergence holds.

1.2.3 Examples

TABLE 1.1: OBJECTIVE FUNCTIONS TO MINIMIZE

Estimator	$Q_n(\theta)$	$Q(\theta)$
MLE [1]	$-\frac{1}{n}\sum_i \log f(W_i, \theta)$	$-\mathbb{E}[\log f(W_i(\theta))]$
Least Squares [2]	$\frac{1}{n}\sum_i (Y_i - g(X_i, \theta))^2/2$	$\mathbb{E}[Y_i - g(X_i, \theta)]^2/2$
GMM [3]	$\|A_n \frac{1}{n}\sum_i g(W_i, \theta)\|^2/2$	$\|A\,\mathbb{E}[g(W_i, \theta)]\|^2/2$
Two-Step [4]	$\|A_n G_n(\theta, \hat{\tau}_n)\|^2/2$	
Minimum Distance [5]	$\|A_n(\hat{\pi}_n - g(\theta))\|^2/2$	$\|A(\pi_0 - g(\theta))\|^2/2$

[1] $f(W, \theta)$ is data density.
[2] $Y_i = g(X_i; \theta_0) + U_i$, where $\mathbb{E}[U_i|X_i] = 0$. For linear least squares, $g(X_i; \theta) = X_i'\theta$.
[3] $\mathbb{E}[g(W_i, \theta_0)] = 0$; A_n is a $(k \times k)$ weighting matrix, where $k \geq d$ (d is the dimension of the parameter θ). The objective function can be thought of as the (weighted) sum of squares of moments.
[4] $G_n(\theta, \tau) \approx 0$ if $\theta = \theta_0$ and $\tau = \tau_0$. Note that $\hat{\tau}_n$ is a consistent estimator of τ_0 that is estimated in the first stage; final standard errors (after the second stage) must account for first stage estimation e.g. weights calculated from a sample in a weighted least squares estimation.
[5] $\hat{\pi}_n$ is a consistent estimator of π_0, which is a k-dimensional vector, $k \geq d$. π_0 is a *known* function (g) of true parameter θ_0: $\pi_0 = g(\theta_0)$. We can think of θ as *structural* parameters, and think of π as *reduced-form* parameters.

[2]To get Assumption 1.2.3, we need Law of Large Numbers (LLN), which holds if we have (i) $\{W_i\}_{i=1}^n$ i.i.d., and (ii) $\|\mathbb{E}[m(W_i, \theta)]\| \leq \infty$ for all $\theta \in \Theta$. Then, $\frac{1}{n}\sum_{i=1}^n m(W_i, \theta) \to^p \mathbb{E}[m(W_i, \theta)]$ for all $\theta \in \Theta$.

1.2.3.A Additional Comments on GMM

Since $g(W_i, \theta) \in \mathbb{R}^k$ for some $k \geq d$, $\mathbb{E}[g(W_i, \theta)] = 0$ gives us a system of k-equations in d-unknowns. For example, we could have

$$\mathbb{E}[g(W_i, \theta)] = \begin{bmatrix} \mathbb{E}[g_1(W_i, \theta)] = 0 \\ \mathbb{E}[g_2(W_i, \theta)] = 0 \end{bmatrix}, \; A_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then,

$$Q_n(\theta) = \frac{1}{2}\left[\left(\frac{1}{n}\sum_{i=1}^{n} g_1(W_i, \theta)\right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n} g_2(W_i, \theta)\right)^2\right]$$

Thus, the objective function can be interpreted as the sum of squares of moment conditions (weighted by the A matrix).

Other ways to write the objective function from the table above include:

1. $Q(\theta) = [\mathbb{E}(g)]' A \, \mathbb{E}(g)$
2. $Q_n(\theta) = [\mathbb{E}_n(g)]' A_n \, \mathbb{E}_n(g)$

where $\mathbb{E}_n(g) = \frac{1}{n}\sum_{i=1}^{n} g$

1.2.4 (Point) Identification

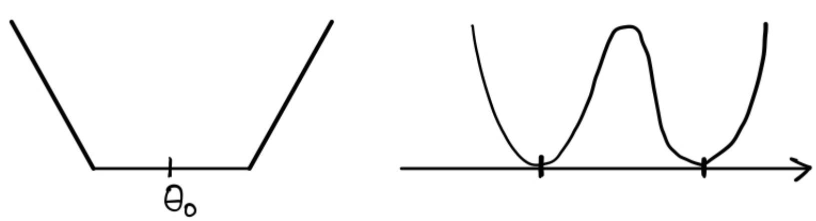
This is a *substantive* assumption. Let $B(\theta, \epsilon)$ be an open ball in Θ of radius ϵ centered at θ.

Assumption 1.2.4 (ID)**.** *There exists a $\theta_0 \in \Theta$ such that for all $\epsilon > 0$,*

$$\inf_{\theta \notin B(\theta_0, \epsilon)} Q(\theta) > Q(\theta_0)$$

In words, we are assuming that if I were to go outside of *any* ϵ-ball centered at true θ_0, the objective function will **strictly** increase. Assumption 1.2.4 is on the **population** objective function $Q(\theta)$ (no "n" here). A necessary condition for identification is that the objective function is uniquely minimized over Θ. So, the objective function cannot look like the functions given in Fig. 1.1

FIGURE 1.1: OBJECTIVE FUNCTION NOT UNIQUELY MINIMIZED



Sufficient conditions are given in Lemma 1.2.

1.2.4.A Examples of Identification

1. **Maximum Likelihood**: Let $f(W, \theta_0)$ be the true density. Then we have

$$\begin{aligned}
Q(\theta_0) - Q(\theta) &= \mathbb{E}[-\log f(W,\theta_0) + \log f(W,\theta)] \\
&= \mathbb{E}[\log(f(W,\theta)/f(W,\theta_0))] \\
&\leq \log \mathbb{E}[f(W,\theta)/f(W,\theta_0)] && \text{(Jensen's Inequality)} \\
&= \log \int \frac{f(W,\theta)}{\cancel{f(W,\theta_0)}} \cancel{f(W,\theta_0)} d\omega = \log(1) = 0 && \text{(Expectation is w.r.t true density)}
\end{aligned}$$

Thus, $Q(\theta_0) - Q(\theta) \leq 0$ and this inequality is strict if $\Pr[f(W,\theta) \neq f(W,\theta_0)] > 0$ for all $\theta \neq \theta_0$. This is the same condition as required for *point identification* of ML.

- **Misspecification:** When the true distribution of the data $f(W)$ is not in $\{f(W,\theta) : \theta \in \Theta\}$, the ML estimator converges in probability to θ_0 that uniquely minimizes the Kullback-Leibler (KL) divergence/information number, $K(f, f(\cdot,\theta))$, between the true density f and the densities in $\{f(\cdot,\theta)\}_{\theta\in\Theta}$, where $K(f, f(\cdot,\theta)) = \mathbb{E}[\log f(W)] - \mathbb{E}[\log f(\cdot,\theta)]$.

2. **Least Squares**: Given $Y = g(X,\theta_0) + U$ and $\mathbb{E}[U|X] = 0$,

$$\begin{aligned}
Q(\theta) - Q(\theta_0) &= \mathbb{E}[(Y - g(X,\theta))^2 - (Y - g(X,\theta_0))^2]/2 \\
&= \mathbb{E}[(g(X,\theta_0) + U - g(X,\theta))^2 - (U)^2]/2 \\
&= \mathbb{E}[(g(X,\theta_0) - g(X,\theta))^2 + 2U(g(X,\theta_0) - g(X,\theta)) + U^2 - U^2]/2 \\
&= \mathbb{E}[(g(X,\theta_0) - g(X,\theta))^2]/2 + 2\,\mathbb{E}[\mathbb{E}[U|X](g(X,\theta_0) - g(X,\theta))]/2 \\
&= \mathbb{E}[(g(X,\theta_0) - g(X,\theta))^2]/2 \geq 0
\end{aligned}$$

The above inequality is strict if and only if $\Pr[g(X,\theta) \neq g(X,\theta_0)] > 0$ for all $\theta \neq \theta_0$.

- **Misspecification:** Suppose true conditional expectation is given by $\mathbb{E}[Y|X] = g(X)$ which is not in the family paramterized by $\theta \in \Theta$. Then,

$$\begin{aligned}
Q(\theta) &= \mathbb{E}[(Y - g(X,\theta))^2]/2 \\
&= \mathbb{E}[(Y - g(X) + g(X) - g(X,\theta))^2]/2 \\
&= \mathbb{E}[(Y - g(X))^2]/2 + \mathbb{E}[(g(X) - g(X,\theta))^2]/2 + 2\,\mathbb{E}[(Y - g(X))(g(X) - g(X,\theta))]/2 \\
&= \mathbb{E}[U^2]/2 + \mathbb{E}[(g(X) - g(X,\theta))^2]/2 + 2\,\underbrace{\mathbb{E}[U(g(X) - g(X,\theta))]}_{=0 \text{ as above}}/2
\end{aligned}$$

Now, the estimated θ uniquely minimizes $\mathbb{E}[(g(X) - g(X,\theta))^2]$. This is the best mean-squared error approximation in the family $\{g(X,\theta) : \theta \in \Theta\}$ to the conditional expectation function.

3. **GMM**: By definition of identification, θ_0 is identified if $\mathbb{E}[g(W,\theta_0)] = 0$ **uniquely** at $\theta = \theta_0$ (The moment condition is not equal to zero at any other θ). For identification, we also need the weighting matrix A to be nonsingular.

4. **Minimum Distance**: If there exists a unique θ_0 such that $\pi_0 = g(\theta_0)$ then identification assumption holds. Recall that $g(\cdot)$ is a known mapping from "structural" parameters θ to π, and minimum distance estimator minimizes the objective function $Q_n(\theta) = -[\hat{\pi}_n - g(\theta)]' A_n [\hat{\pi}_n - g(\theta)]/2$

5. **Two-Step Estimators**: Assumption 1.2.4 satisfied if there exists a unique θ_0 such that $G(\theta_0, \tau_0) = 0$.

6. **Probit**: Setup (i) $W = (Y, X)$, (ii) $Y \in \{0, 1\}$, (iii) $X \in \mathbb{R}^k$, (iv) $Y = \mathbb{1}\{\epsilon \leq X'\theta\}$ (v) $\epsilon \perp X_i$ (vi) $\epsilon \sim N(0, 1)$ (vii) $f^*(W) = f(W; \theta_0) = \Phi(X'\theta_0)^Y [1 - \Phi(X'\theta_0)]^{1-Y}$

   To prove that identification holds, we need to show that $\forall \theta \neq \theta_0$, $\exists X$ s.t. $X'\theta \neq X'\theta_0$.

   Claim: $\mathbb{E}[XX']$ exists and is nonsingular $\Rightarrow \theta_0$ uniquely minimizes $Q(\theta)$.

   Proof. Nonsingular $\mathbb{E}[XX']$ implies that it is positive definite. That is, for $\theta \neq \theta_0$,

$$
\begin{aligned}
\mathbb{E}[\{X'(\theta - \theta_0)\}^2] &= (\theta - \theta_0)' \, \mathbb{E}[XX'](\theta - \theta_0) > 0 \\
&\Rightarrow X'(\theta - \theta_0) \neq 0 \\
&\Rightarrow X'\theta \neq X'\theta_0
\end{aligned}
$$

   where the last inequality simply means that $X'\theta$ and $X'\theta_0$ are *not equal on a set of positive probability* – that is, there exists some X (with Pr $= 1$) such that $X'\theta \neq X'\theta_0$ (if it was equal everywhere, the expectation will be zero). Both $\Phi(X'\theta)$ and $\Phi(-X'\theta)$ are strictly monotonic, so that $X'\theta \neq X'\theta_0$ implies both $\Phi(X'\theta) \neq \Phi(X'\theta_0)$ and $1 - \Phi(X'\theta) \neq 1 - \Phi(X'\theta_0)$, and hence $f(W; \theta) \neq f(W; \theta_0)$ ■

1.2.4.B Primitive Conditions for Identification

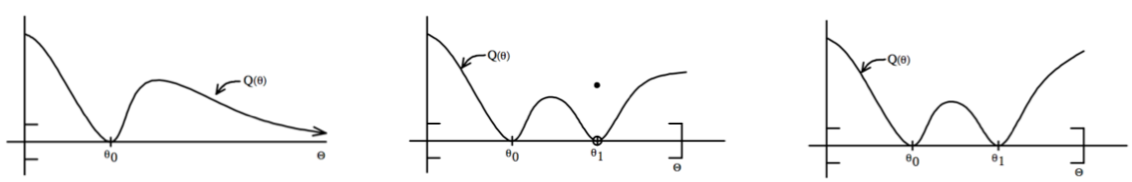
Lemma 1.2

Let

1. Θ be compact.
2. $Q(\theta)$ continuous, and
3. θ_0 uniquely minimizes $Q(\theta)$ over $\theta \in \Theta$

Then, identification (Assumption 1.2.4) holds.

FIGURE 1.2: ALL THREE CONDITIONS MUST BE SATISFIED



From left to right, violation of compactness, continuity, and uniqueness.

In the middle figure, the inf condition is not satisfied at θ_1.

1.2.5 Main Result

Theorem 1.1: Consistency

Assumption 1.2.1 (**EE**), Assumption 1.2.2 (**UCONV**), and Assumption 1.2.4 (**ID**) together imply consistency:

$$\hat{\theta}_n \to^p \theta_0$$

Proof. From Assumption 1.2.4, for a given $\epsilon > 0$, $\exists \delta > 0$ such that

$$\theta \notin B(\theta_0, \epsilon) \Rightarrow Q(\theta) - Q(\theta_0) \geq \delta > 0$$

Then,

$$\begin{aligned}
\Pr\Big[\hat{\theta}_n \notin B(\theta_0, \epsilon)\Big] &\leq \Pr\Big[Q(\hat{\theta}_n) - Q(\theta_0) \geq \delta\Big] \\
&= \Pr\Big[Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\hat{\theta}_n) - Q(\theta_0) \geq \delta\Big] \\
&\leq \Pr\Big[Q(\hat{\theta}_n) - Q_n(\hat{\theta}_n) + Q_n(\theta_0) + o_p(1) - Q(\theta_0) \geq \delta\Big] && \text{(Assumption 1.2.1)} \\
&\leq \Pr\left[2 \sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \geq \delta\right] \to^p 0 && \text{(Assumption 1.2.2)}
\end{aligned}$$

■

1.2.5.A Extra: Primitive Conditions for Continuity of an Expectation

$\mathbb{E}[m(W, \theta)]$ continuous if

1. $m(W, \theta)$ continuous in θ for all W
2. $\mathbb{E}[\sup_\theta \|m(W, \theta)\|] < \infty$

*Continuity is a weak assumption since kinks are allowed.

1.3 Normality

Chapter 2

Introduction

Hello, here is some text without a meaning. This...