Properties of Extremum Estimators

Asymptotic Theory – Part III

James J. Heckman
University of Chicago

Econ 312
This draft, April 12, 2006
As we saw in an earlier lecture (Asymptotic Theory – Part II), the Maximum Likelihood Estimator, Nonlinear Least Squares Estimator (NLS) and even the OLS estimator are all examples of “Extremum Estimators”. In this lecture we examine theorems and proofs for the consistency and asymptotic normality of Extremum Estimators in a somewhat specialized, but easily generalized, form.
The following theorems lay out the conditions under which extremum estimators are consistent and asymptotically normal. They each talk about estimators using the maximum principle, but can trivially be extended to minimum principle estimators by placing a negative sign in front of $Q(y, \theta)$,\(^1\) i.e. $\min Q = \max[-Q]$.

\(^1\)In this lecture, \(\{y\}\) denotes all the data, and hence includes both dependent and independent variables (corresponding to \(\{x\}\) and \(\{y\}\) in the earlier lecture).
1 Consistency of extremum estimators

The first theorem proves consistency when the criterion function has a globally unique maximum or minimum, respectively in the population. Thus $\Theta$ is uniquely identified. Differentiability of $Q_T(\theta)$ is not required.

The second theorem states the additional assumptions you have to make if $Q$ is only locally identified, i.e. there are multiple solutions to $\{\max Q\}$ but only one is in the neighborhood $N(\theta_0)$ of $\theta_0$. It assumes differentiability of $Q_T(\theta)$. 
Theorem 1 (Global): Assume that

1. Parameter space $\Theta$ is a compact subset of $R^K$;

2. $Q_T(y, \theta)$ is continuous in $\theta, \theta \in \Theta, \forall y$ and is a measurable function of $y, \forall \theta \in \Theta$;

3. $Q_T(y, \theta) \xrightarrow{Pr} Q(\theta)$, a nonstochastic function, in probability uniformly as $T \to \infty$; and

4. $\theta_0 = \arg \max_{\theta \in \Theta} Q(\theta)$ is globally identified. (i.e. $Q(\theta)$ achieves global maximum at $\theta_0$).

If we let $\hat{\theta}_T = \arg \max_{\theta \in \Theta} Q_T (y, \theta)$, then: $\hat{\theta}_T \xrightarrow{Pr} \theta_0$. 

Observe that continuity of $Q(\theta)$ follows from the fact that limits of uniformly continuous functions are continuous, and continuity of $Q$ in $\theta$ and compactness of $\Theta$ implies uniform continuity of $Q(\theta)$. 
Proof. Let $N(\theta_0)$ be an open neighborhood in $R^k$ containing $\theta_0$. Then $N^c(\theta_0)$, the complement of $N(\theta_0)$, is closed, so $B \equiv N^c(\theta_0) \cap \Theta$ is compact.

$\Rightarrow \max_{\theta \in B} Q(\theta)$ exists.

Denote $\varepsilon = [Q(\theta_0) - \max_{\theta \in B} Q(\theta)] > 0$.

Let $A_T$ be the event

$$A_T = \{|Q_T(\theta) - Q(\theta)| < \varepsilon/2\}$$
$$= \{-\varepsilon/2 < Q_T(\theta) - Q(\theta) < \varepsilon/2\} \quad \forall \theta \in \Theta.$$

This event is “likely” with $T$ big due to assumption (3) (uniform convergence of $Q_T$ to $Q$), i.e.:

$$Q_T \xrightarrow{\text{pr. uniformly}} Q \quad \Rightarrow \quad \Pr \{A_T\} \to 1 \text{ as } T \to \infty \quad (*)$$
Then $A_T$ implies:

1. $Q\left(\hat{\theta}_T\right) > Q_T\left(\hat{\theta}_T\right) - \varepsilon/2$

2. $Q_T\left(\theta_0\right) > Q\left(\theta_0\right) - \varepsilon/2$

Also we have $Q_T\left(\hat{\theta}_T\right) > Q_T\left(\theta_0\right)$ by the definition of $\hat{\theta}_T$. Then from the above facts we get:

$$Q(\hat{\theta}_T) > Q_T(\hat{\theta}_T) - \varepsilon/2 > Q_T(\theta_0) - \varepsilon/2 > Q(\theta_0) - \varepsilon.$$

$\therefore Q(\hat{\theta}_T) > Q(\theta_0) - \varepsilon$. Since we have a strict inequality, from the definition of $\varepsilon$, we get that:

$$A_T \Rightarrow \{\hat{\theta}_T \in N(\theta_0)\}, \text{ for } T \text{ sufficiently large.}$$
Then it must be that:

\[ \Pr\{A_T\} \leq \Pr\{\hat{\theta}_T \in N(\theta_0)\}. \]

Then, from equation (*) we have that:

\[
\lim_{T \to \infty} \Pr\{A_T\} = 1 \Rightarrow \lim_{T \to \infty} \Pr\{\hat{\theta}_T \in N(\theta_0)\} = 1
\]

and so \( \hat{\theta}_T \xrightarrow{Pr} \theta_0 \), because choice of \( \varepsilon \) is arbitrary. \( \blacksquare \)
Theorem 2 (Local): Assume that:

1. Parameter space $\Theta$ is an open subset of $\mathbb{R}^K$ that contains $\theta_0$;

2. $Q_T(y, \theta)$ is a measurable function of $y \; \forall \theta \in \Theta$;

3. $\frac{\partial Q_T}{\partial \theta}$ exists and is continuous in an open neighborhood $N_1(\theta_0)$ of $\theta_0$ (this implies $Q_T$ is continuous $\forall \theta \in N_1(\theta_0)$);

4. There exists an open neighborhood $N_2(\theta_0)$ of $\theta_0$ such that $Q_T(y, \theta) \to Q(\theta)$, a non-stochastic function, in probability uniformly $\forall \theta \in N_2(\theta_0)$ as $T \to \infty$; and

5. $\theta_0 = \arg\max_{\theta \in N_2(\theta_0)} Q(\theta)$ is locally identified.
If we let $\hat{\Theta}_T$ denote the set of roots of $\frac{\partial Q_T}{\partial \theta} = 0$ corresponding to the local maxima; then, for any $\varepsilon > 0$,

$$\lim_{T \to \infty} \Pr \left\{ \theta \in \hat{\Theta}_T \mid \inf |\theta - \theta_0| > 0 \right\} = 0.$$ 

Proof. See Amemiya, chapter 4. □
2 Asymptotic normality of extremum estimators

Now we will show that under certain conditions on the first and second derivatives of $Q$, the criterion function for an estimator which uses the extremum principle, the asymptotic distribution of the extremum estimator $\hat{\theta}_T$ (chosen as the maximizer of $Q_T$) is normal.
Theorem 3 (Cramer): Assume the conditions of Theorem 2, in addition:

1. $\frac{\partial^2 Q_T}{\partial \theta \partial \theta'}$ exists and is continuous in an open neighborhood of $\theta_0$;

2. There exists an open neighborhood $N(\theta_0)$ of $\theta_0$ such that $Q_T(y, \theta) \to Q(\theta)$, a nonstochastic function, in probability uniformly $\forall \theta \in N(\theta_0)$ as $T \to \infty$.

3. $\left. \frac{\partial^2 Q_T(\theta)}{\partial \theta \partial \theta'} \right|_{\theta^*_T} \overset{p}{\to} A(\theta_0)$ if $\theta^*_T \overset{p}{\to} \theta_0$, where

$$A(\theta_0) = p \lim \left( \frac{\partial^2 Q_T(\theta)}{\partial \theta \partial \theta'} \right)_{\theta_0}$$

is nonsingular; and
4. $\sqrt{T} \left( \frac{\partial Q_T(\theta)}{\partial \theta} \right) \bigg|_{\theta_0} \sim N(0, B(\theta_0))$, where

$B(\theta_0) = E \left[ \frac{\partial Q_T(\theta)}{\partial \theta} \cdot \frac{\partial Q_T(\theta)'}{\partial \theta} \right]_{\theta_0}$.

If we let $\hat{\theta}_T$ denote the root of $\frac{\partial Q_T}{\partial \theta} = 0$, then:

$\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \sim N(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1})$. 
**Proof.** By assumption we have: \( \frac{\partial Q_T}{\partial \theta} \bigg|_{\hat{\theta}_T} = 0 \). Then taking a Taylor expansion of the l.h.s. around \( \theta_0 \), we have

\[
\frac{\partial Q_T}{\partial \theta} \bigg|_{\hat{\theta}_T} = \frac{\partial Q_T}{\partial \theta} \bigg|_{\theta_0} + \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \bigg|_{\theta^*} \left( \hat{\theta}_T - \theta_0 \right) + o_P(1),
\]

where \( \theta^* \) lies between \( \hat{\theta}_T \) and \( \theta_0 \). Multiplying by \( \sqrt{T} \), we get:

\[
o_P(1) + \sqrt{T} \frac{\partial Q_T}{\partial \theta} \bigg|_{\theta_0} + \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \bigg|_{\theta^*} \sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) = 0.
\]

Rearranging, we get:

\[
\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) = - \left( \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \bigg|_{\theta^*} \right)^{-1} \sqrt{T} \frac{\partial Q_T}{\partial \theta} \bigg|_{\theta_0} + o_P(1).
\]
Since $\hat{\theta}_T \xrightarrow{p} \theta_0 \implies \theta^* \xrightarrow{p} \theta_0$, we see the first object on the r.h.s. becomes:

$$\left. \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \right|_{\theta^*} \xrightarrow{p} \left. \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \right|_{\theta_0} = A(\theta_0),$$

where $A(\theta_0)$ is constant. As for the second object on the r.h.s., by assumption,

$$\sqrt{T} \left( \frac{\partial Q_T(\theta)}{\partial \theta} \right|_{\theta_0} \right) \sim N \left( 0, B(\theta_0) \right).$$

Putting this all together we have, by Slutsky’s Theorem,

$$\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \sim N \left( 0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0) \right).$$
Observe that assumption (4) is a consequence of a uniform central limit theorem.

\[ \sqrt{T} \left( \frac{\partial Q_T(\theta)}{\partial \theta} \bigg|_{\theta_0} \right) = \frac{1}{\sqrt{T}} \left( \sum_{t=1}^{T} q_t(\theta) \right), \]

i.i.d. random variables with mean zero and we norm them by \( \sqrt{T} \). We get, by a CLT, that the variance of this random variable is

\[ E \left[ \left( \frac{\partial Q_T(\theta)}{\partial \theta} \right) \cdot \left( \frac{\partial Q_T(\theta)}{\partial \theta} \right)' \right]. \]
References
