Economics 481
Solutions to Final Examination

Oliver Linton

March 6, 2000

1. Let Z_i, $i = 1, \ldots, n$ be i.i.d. and define

$$T_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i, \theta) - Em(Z_i, \theta),$$

where θ is a parameter and m is some given function. A pointwise weak law of large numbers would assert that for any given θ we have

$$T_n(\theta) \to_p 0,$$

while a strong law would have the convergence in probability replaced by convergence almost surely. A uniform law of large numbers on the other hand would assert that

$$\sup_{\theta \in \Theta} |T_n(\theta)| \to_p 0,$$

where the set Θ.

The importance of this result is that it allows us to replace the sample moment condition $G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i, \theta)$ by the population moment $G(\theta) = Em(Z_i, \theta)$. We want to argue that we can replace the minimizer of $\|G_n(\theta)\|$, which is $\widehat{\theta}$, by the minimizer of $\|G(\theta)\|$, which is θ_0, with high probability. Pointwise convergence is not enough as can be seen from the deterministic example

$$\|G_n(\theta)\| = \begin{cases} 1 - \frac{\theta^2}{\theta^2 + (1-n\theta)^2} & 0 \le \theta < 1 \\ \frac{1}{2} & \theta = 1 \end{cases} \quad , \quad n = 1, 2, \ldots$$

Then,

$$\lim_{n\to\infty} \|G_n(\theta)\| = \|G(\theta)\| = \begin{cases} 1 & 0 \le \theta < 1 \\ \frac{1}{2} & \theta = 1. \end{cases}$$

However, the minimizer of $\|G_n(\theta)\|$ is $\theta = 1/n$ for all n, which gives $\|G_n(\theta)\| = 0$, while the minimizer of $\|G(\theta)\|$ is $\theta = 1$ which gives $\|G(\theta)\| = 1/2$.

Let x_{jk} be the value of x that satisfies $F(x_{jk}) = j/k$ for integer j, k with $j \leq k$. For any x between x_{jk} and $x_{j+1,k}$,

$$F(x_{jk}) \leq F(x) \leq F(x_{j+1,k}) \quad ; \quad F_n(x_{jk}) \leq F_n(x) \leq F_n(x_{j+1,k}),$$

while $0 \leq F(x_{j+1,k}) - F(x_{jk}) \leq 1/k$, so that

$$\begin{aligned}
F_n(x) - F(x) &\leq F_n(x_{j+1,k}) - F(x_{jk}) \leq F_n(x_{j+1,k}) - F(x_{j+1,k}) + \frac{1}{k} \\
F_n(x) - F(x) &\geq F_n(x_{j,k}) - F(x_{j+1,k}) \geq F_n(x_{j,k}) - F(x_{j,k}) - \frac{1}{k}.
\end{aligned}$$

Therefore, for any x and k,

$$|F_n(x) - F(x)| \leq \max_{1 \leq j \leq k} |F_n(x_{jk}) - F(x_{jk})| + \frac{1}{k}. \tag{1}$$

Since the right hand side of (1) does not depend on x, we can replace the left hand side by $\sup_{-\infty < x < \infty} |F_n(x) - F(x)|$. We take $k = \log n$, which ensures that

$$\sup_x |F_n(x) - F(x)| \leq \max_{1 \leq j \leq k} |F_n(x_{jk}) - F(x_{jk})| + o(1)$$

as $n \to \infty$. For any $\epsilon > 0$, let $A_j = \{|F_n(x_{jk}) - F(x_{jk})| > \epsilon\}$ and

$$A = \cup_{j=1}^k A_j = \{\max_{1 \leq j \leq k} |F_n(x_{jk}) - F(x_{jk})| > \epsilon\}.$$

Then

$$\begin{aligned}
\Pr(A) &\leq \sum_{j=1}^k \Pr(A_j) \\
&\leq \sum_{j=1}^k \frac{E\left(|F_n(x_{jk}) - F(x_{jk})|^2\right)}{\epsilon^2} \\
&= \sum_{j=1}^k \frac{F(x_{jk})(1 - F(x_{jk}))}{\epsilon^2 n} \\
&= O(\log n/n).
\end{aligned}$$

The first inequality follows by the Bonferroni inequality, while the second one uses the Chebychev or Markov inequality. ■

2. We have

$$\begin{aligned} f_A(x) &= \frac{d}{dx}\Pr\left[A \le x\right] \\ &= \frac{d}{dx} E \Pr\left[L \le \frac{x}{W}\right] \\ &= \frac{d}{dx}\int_0^\infty F_L(\frac{x}{z}) f_W(z)dz \\ & \int_0^\infty f_L(\frac{x}{z})\frac{1}{z} f_W(z)dz \end{aligned}$$

as required.

Note that

$$\partial l_n/\partial\theta = \sum_{i=1}^n \frac{\partial \log f_A}{\partial \theta}(A_i;\theta) = \sum_{i=1}^n \frac{1}{f_A}\frac{\partial f_A}{\partial \theta}(A_i;\theta),$$

where

$$\begin{aligned} \frac{\partial f_A}{\partial L}(A_i;\theta) &= \int_0^\infty \frac{\partial f_L}{\partial L_0}(\frac{A_i}{z};L)\frac{1}{z} f_W(z;W)dz \\ \frac{\partial f_A}{\partial W}(A_i;\theta) &= \int_0^\infty f_L(\frac{A_i}{z};L)\frac{1}{z}\frac{\partial f_W}{\partial W}(z;W)dz \end{aligned}$$

Then we can obtain an unbiased simulator of these two derivatives as follows. First rewrite the first derivatives as

$$\frac{\partial f_A}{\partial L}(A_i;\theta) = \int_0^\infty h_i(z;\theta) f_1(z)dz, \quad h_i(z;\theta) = \frac{\frac{\partial f_L}{\partial L}(\frac{A_i}{z};L) f_W(z;W)}{z f_1(z)},$$

and where f_1 is the density of a chi-squared random variable with one degree of freedom. Then let $\varepsilon_j \sim \chi^2(1)$, $j = 1, \ldots, ns$, and let

$$\frac{\partial f_A^{ns}}{\partial L}(A_i;\theta) = \frac{1}{ns}\sum_{j=1}^{ns} h_i(\varepsilon_j;\theta).$$

Similarly,

$$f_A^{ns}(A_i;\theta) = \frac{1}{ns}\sum_{j=1}^{ns} g_i(\varepsilon_j;\theta), \quad g_i(z;\theta) = \frac{f_L(\frac{A_i}{z};L) f_W(z;W)}{z f_1(z)}$$

is unbiased. However,

$$E_{sim}\frac{1}{f_A^{ns}(A_i;\theta)} \neq \frac{1}{f_A(A_i;\theta)}$$

in general, and so even if we use independent samples to compute $f_A^{ns}(A_i;\theta)$ and the two derivative estimators,

$$\begin{aligned}
E_{sim}\frac{1}{f_A^{ns,1}(A_i;\theta)}\frac{\partial f_A^{ns,2}}{\partial\theta}(A_i;\theta) &= E_{sim}\frac{1}{f_A^{ns,1}(A_i;\theta)}E_{sim}\frac{\partial f_A^{ns,2}}{\partial\theta}(A_i;\theta) \\
&= \frac{\partial f_A}{\partial\theta}(A_i;\theta)E_{sim}\frac{1}{f_A^{ns,1}(A_i;\theta)} \\
&\neq \frac{1}{f_A(A_i;\theta)}\frac{\partial f_A}{\partial\theta}(A_i;\theta)
\end{aligned}$$

as would be required for an unbiased simulator of the likelihood score function. Unless we can take $ns \to \infty$ we will generally have problems establishing consistency for this procedure.

By contrast the method of moments is quite easy here. We can take either the following two moments

$$\begin{aligned}
m_1(A_i,\theta) &= A_i - WL \\
m_2(A_i,\theta) &= (A_i - WL)^2 - 2WL
\end{aligned}$$

or

$$\begin{aligned}
m_1(A_i,\theta) &= A_i - WL \\
m_2(A_i,\theta) &= A_i^2 - 2WL - W^2L^2
\end{aligned}$$

should give consistent and asymptotic normal estimators.

■

3.

Suppose you are interested in estimating a parameter vector $\theta = (\theta_1, \theta_2)$ from a sample moment condition $G_n(\theta)$ [for which $G(\theta) = EG_n(\theta)$ is equal to zero if and only if $\theta = \theta_0$], and that a preliminary estimator $\widehat{\theta}_2$ of θ_2 is available. Outline what conditions are needed for $\widehat{\theta}_2$ to satisfy for the estimator $\widehat{\theta}_1$ that sets $G_n(\theta_1, \widehat{\theta}_2) = 0$ to be consistent and asymptotically normal. Discuss the circumstances under which the distribution of $\widehat{\theta}_2$ affects the distribution of $\widehat{\theta}_1$.

■

4. This is the class of estimators that minimizes $\|G_n(\theta)\|_W^2 = G_n(\theta)'WG_n(\theta)$ with respect to θ, where W is a 2 by 2 weighting matrix and

$$G_n(\theta) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (X_i - \theta) \\ \frac{1}{n}\sum_{i=1}^n (X_i - \theta)^3 \end{pmatrix}.$$

To determine the asymptotic distribution we just need to compute

$$\Gamma = \frac{\partial}{\partial\theta} E G_n(\theta_0) \text{ and } V = Emm'.$$

We have

$$V = \begin{pmatrix} E(X_i - \theta_0)^2 & E(X_i - \theta_0)^4 \\ E(X_i - \theta_0)^4 & E(X_i - \theta_0)^6 \end{pmatrix},$$

while

$$\Gamma = E \begin{pmatrix} -1 \\ -3(X_i - \theta_0)^2 \end{pmatrix} = -\begin{pmatrix} 1 \\ 3\sigma^2 \end{pmatrix}.$$

The estimator is asymptotically normal with mean zero and variance covariance matrix

$$(\Gamma' W \Gamma)^{-1} \Gamma' W V W \Gamma (\Gamma' W \Gamma)^{-1}.$$

The optimal weighting matrix is $W_{opt} = V^{-1}$ in which case the asymptotic variance becomes

$$(\Gamma' V^{-1} \Gamma)^{-1}.$$

When X_i is normally distributed, the expression for V simplifies so that

$$V = \begin{pmatrix} \sigma^2 & 3\sigma^4 \\ 3\sigma^4 & 15\sigma^6 \end{pmatrix}.$$

Now,

$$V^{-1} = \begin{pmatrix} \frac{5}{2\sigma^2} & -\frac{1}{2\sigma^4} \\ -\frac{1}{2\sigma^4} & \frac{1}{6\sigma^6} \end{pmatrix},$$

which means that

$$(\Gamma' V^{-1} \Gamma)^{-1} = \frac{1}{\begin{pmatrix} 1 & 3\sigma^2 \end{pmatrix} \begin{pmatrix} \frac{5}{2\sigma^2} & -\frac{1}{2\sigma^4} \\ -\frac{1}{2\sigma^4} & \frac{1}{6\sigma^6} \end{pmatrix} \begin{pmatrix} 1 \\ 3\sigma^2 \end{pmatrix}} = \sigma^2.$$

The asymptotic variance in this case as the same as the variance of the sample mean, which is known to be the efficient estimator.

The efficient estimator is asymptotically equivalent to a solution to the exactly identified system

$$\alpha \frac{1}{n} \sum_{i=1}^{n} (X_i - \theta) + \beta \frac{1}{n} \sum_{i=1}^{n} (X_i - \theta)^3 = 0,$$

where

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \propto V^{-1} \Gamma.$$

The constant factor of proportional is irrelevant here. In the normal case, we can take $\alpha = 1$ and $\beta = 0$, i.e., the optimal combination of the moments is just to take the first moment.

More generally, both moments will be used and $\beta \neq 0$. Since the variance of the sample mean is σ^2 regardless of the distributional shape, it must be that the variance of the optimal GMM estimator is less than or equal to σ^2. ■

5. We use the general theorem

Theorem 1 *Assume* $\|G(\theta_0)\| = 0$, *and*

1. *Define Estimator.*
$$\|G_n(\widehat{\theta})\| = \inf_{\theta\in\Theta} \|G_n(\theta)\| + o_p(1),$$

2. *Identification.*
$$\text{for all } \delta > 0, \quad \inf_{\|\theta-\theta_0\|>\delta} \|G(\theta)\| \geq \epsilon(\delta) > 0,$$

3. *ULLN.*
$$\sup_{\theta\in\Theta} \|G_n(\theta) - G(\theta)\| = o_p(1).$$

Then
$$\widehat{\theta} - \theta_0 = o_p(1).$$

By the law of large numbers [which holds because the random variables are bounded]

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} H_\alpha(y_i - \theta) &\rightarrow {}_pEH_\alpha(y_i - \theta) \\
&= E\text{sign}(y_i - \theta) + (2\alpha - 1) \\
&= \Pr(y_i > \theta) - \Pr(y_i \leq \theta) + (2\alpha - 1) \\
&= 1 - 2F_y(\theta) + (2\alpha - 1) \equiv G(\theta),
\end{aligned}$$

where F_y is the c.d.f. of y_i.

This convergence is uniform in θ because $\{y_i > \theta\}$ is continuous in θ with probability one since y is continuously distributed.

Now

$$G(\theta_0(\alpha)) = 1 - 2F_y(\theta_0(\alpha)) + 2\alpha - 1 = 1 - 2\alpha + 2\alpha - 1 = 0.$$

Furthermore, the function G is monotonic in θ provided F_y is strictly monotonic at $\theta = \theta_0(\alpha)$.

Under additional assumptions, we can show that $\widehat{\theta}$ is asymptotically normal with mean zero and variance V/Γ^2, where $V = Em^2$ and $\Gamma = G'(\theta)$. Here,

$$\begin{aligned} V &= EH_\alpha^2(y_i - \theta_0(\alpha)) \\ &= E[1 + 2\text{sign}(y_i - \theta_0(\alpha))(2\alpha - 1) + (2\alpha - 1)^2\}] \\ &= 1 + 2(1 - 2\alpha)(2\alpha - 1) + (2\alpha - 1)^2 \\ &= 4\alpha(1 - \alpha), \end{aligned}$$

while

$$\Gamma = -2f_y(\theta_0(\alpha)).$$

In conclusion, the asymptotic variance of the estimator is $\alpha(1-\alpha)/f_y^2(\theta_0(\alpha))$.

We now need a first order condition for each parameter β, θ. One suitable moment condition is

$$\frac{1}{n}\sum_{i=1}^n w_i H_\alpha(x_i - \theta - \beta z_i), \quad w_i = \begin{pmatrix} 1 \\ z_i \end{pmatrix}.$$

This will be consistent. ■

6. To show that $\widehat{\beta}$ is consistent we need to show that uniformly in $\beta \in B$

$$G_n(\beta) \to_p G(\beta)$$

for some fixed function G that is uniquely zeroed at $\beta = \beta_0$. Write

$$G_n(\beta) = \frac{1}{n}\sum_{t=1}^n \varepsilon_t t^\alpha + \frac{1}{n}\sum_{t=1}^n (F(\beta_0 t^\alpha) - F(\beta t^\alpha))t^\alpha.$$

The first term does not depend on β and has mean zero and variance

$$\frac{1}{n^2}\sum_{t=1}^n F(\beta_0 t^\alpha)(1 - F(\beta_0 t^\alpha))t^{2\alpha} \simeq \frac{1}{n^2}\sum_{t=1}^n \frac{1}{\beta_0 t^\alpha} t^{2\alpha} = O(n^{\alpha-1}).$$

This will go zero if and only if $\alpha < 1$.

Secondly,

$$\frac{1}{n}\sum_{t=1}^n (F(\beta_0 t^\alpha) - F(\beta t^\alpha))t^\alpha \simeq \frac{1}{n}\sum_{t=1}^n (1 - \frac{1}{\beta_0 t^\alpha} - (1 - \frac{1}{\beta t^\alpha}))t^\alpha \simeq \frac{1}{\beta} - \frac{1}{\beta_0}$$

as $n \to \infty$. The convergence of this deterministic quantity is uniform in $\beta \in B$ because the left hand side is bounded. Thus we can take $G(\beta) = (\beta_0 - \beta)/\beta\beta_0$ which has a unique zero at $\beta = \beta_0$.

For the asymptotic distribution here we need to find that normalization such that $G_n(\beta_0) = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t t^\alpha$ is asymptotically normal, because

$$\beta - \beta_0 \simeq -\left[\frac{\partial G}{\partial \beta}(\beta_0)\right]^{-1} G_n(\beta_0)$$

We have already calculated that the variance of $G_n(\beta_0)$ is $\frac{1}{n^2}\sum_{t=1}^{n} \frac{1}{\beta_0} t^\alpha = O(n^{\alpha-1})$. In fact, we can show that

$$n^{(1-\alpha)/2} G_n(\beta_0) \to_d N(0, \frac{1}{\beta_0(\alpha+1)}),$$

and so

$$n^{(1-\alpha)/2}(\widehat{\beta} - \beta_0) \to_d N(0, \frac{\beta_0^3}{(\alpha+1)}).$$

If $\alpha = 1$, we have $G_n(\beta_0) = O_p(1)$ so that

$$\beta - \beta_0 \simeq -\left[\frac{\partial G}{\partial \beta}(\beta_0)\right]^{-1} G_n(\beta_0) = O_p(1).$$

■