1 Econometrics II: Homework 5

Due: Thursday, April 17^{th}

1 Consider estimation of a parameter $\theta_0 \in \mathbb{R}$. Assume that the estimator $\hat{\theta}$ is defined as $\hat{\theta} = \arg\min_{\theta\in\mathbb{R}} \frac{1}{n}\sum_m (W_i, \theta)$ for a known, real-valued function $m(\cdot)$ and that the following conditions hold:

(i) $m(w, \theta)$ is strictly convex in θ for every w.
(ii) $\mathbb{E}m(W, \theta)$ exists and is finite for all θ.
(iii) $\theta_0 = \arg\min_{\theta\in\Theta} \mathbb{E}m(W, \theta)$ exists.
(iv) $\{W_i\}$ is i.i.d.
Then it is true that $\hat{\theta} \xrightarrow{p} \theta_0$.

1.1 Where do we assume less than in the similar theorem from class? (Consider in particular the specialization to m-estimators. The question is *not* about maximization versus minimization.) Where do we assume more? Are those assumptions from the lecture notes that are stronger than assumptions above implied?

1.2 Prove the claim. (For simplicity, set $\mathbb{E}m(\theta_0, W) = 0$. This is without loss of generality as we could always replace $m(\cdot)$ with $m(\cdot) - \mathbb{E}m(\theta_0, W)$.)

2 Replicate Worked Example 1 but using the Geometric distribution. This is the distribution of number of trials needed to observe one success with i.i.d. tosses of a coin with parameter π.

2.1 Verify that the r.v. X is characterized by probability mass function and moments

$$\begin{aligned}
\Pr(X = x) &= (1-\pi)^{x-1}\pi \\
\mathbb{E}X &= 1/\pi \\
\mathrm{var}(X) &= (1-\pi)/\pi^2.
\end{aligned}$$

2.2 Find the Maximum Likelihood estimator and characterize its asymptotic distribution with greatest possible precision.

2.3 Is the estimator unbiased? If not, can you sign its bias?

3 We consider estimation of $\mu = \mathbb{E}X$, where $X \sim_{i.i.d.} \mathcal{U}[\alpha, \beta]$, the uniform distribution on the interval $[\alpha, \beta]$. The parameter space is $\mathbb{R}$.

You may use the following facts:

- $\text{var}(X) = (\beta - \alpha)^2/12$.
- If $X_{[k]}$ is the k'th smallest realization ("order statistic") from a sample of size n, then

$$\begin{aligned} \mathbb{E}X_{[k]} &= \alpha + \frac{k}{n+1}(\beta - \alpha), \\ \text{var}(X_{[k]}) &= (\beta - \alpha)^2 \times \frac{k(n+1-k)}{(n+1)^2(n+2)}. \end{aligned}$$

  The correlation between lowest and highest order statistic is $1/n$.

- Convergence in mean squared implies convergence in probability, i.e. if a sequence of scalar-valued random variables $\{Y_n\}$ and a scalar y fulfil $\mathbb{E}((Y_n - y)^2) \to 0$, then $Y_n \stackrel{p}{\to} y$.

3.1 We first derive the usual analog estimator, $\hat{\mu} = \bar{x}_n$ (i.e. the sample average), as a GMM estimator. Define a function $g(x, \mu)$ that achieves this, i.e. (i) $\mathbb{E}g(X, \mu) = 0$ iff $\mu = \mu_0$, (ii) the sample average emerges as corresponding GMM estimator. Without using Theorem 1 below, state the estimator's asymptotic distribution as precisely as possible.

3.2 Next, recover this same asymptotic distribution using Theorem 1 below. Please verify that the theorem's assumption apply and write down expressions for $\boldsymbol{W}$, $\boldsymbol{G}$, and $\boldsymbol{S}$. (The answers should be very simple.)

3.3 Write down the model's likelihood function. (Please start with one observation and use f for likelihood.) Show that the lowest and highest order statistic are the ML estimators $(\hat{\alpha}_{ML}, \hat{\beta}_{ML})$ of α and β.

3.4 Consider the estimator $\tilde{\mu} = (\hat{\alpha}_{ML} + \hat{\beta}_{ML})/2$. Compute bias, variance, and mean-square error (remember: "MSE:=$\mathbb{E}((\hat{\theta} - \theta_0)^2)$ = squared bias + variance") for both $\hat{\mu}$ and $\tilde{\mu}$.

3.5 Does the Gauss-Markov Theorem apply to $\hat{\mu}$? If no, why not? If yes, why is your finding from 4 consistent with it?

3.6 The estimator $\tilde{\mu}$ can be interpreted as an extremum estimator using sample criterion

$$Q_n(\mu) = (x_{[n]} - \mu)^2 + (\mu - x_{[1]})^2.$$

Explain in detail (what is the corresponding Q?) and argue that one consistency theorem for extremum estimators from class applies.

3.7 Describe as precisely as possible the limiting behavior of $\sqrt{n}(\tilde{\mu} - \mu_0)$.

3.8 Consider Theorem 2 below. Your answer to 7 should contradict its conclusion. Why is the theorem not contradicted?

Theorem 1: Asymptotic Distribution of GMM Estimators

Assume that

1. $\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$,
2. $\hat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$, where $\boldsymbol{W}$ is symmetric and positive definite,
3. $\theta_0 \in \operatorname{int}\Theta$,
4. there exists a neighborhood $\mathcal{N}$ of θ_0 s.t. for any $\theta \in \mathcal{N}$, $g(w, \theta)$ is continuously differentiable in θ for any w,
5. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(W, \theta_0) \xrightarrow{d} N(0, \boldsymbol{S})$, $\boldsymbol{S}$ positive definite,
6. $\overline{\theta} \xrightarrow{p} \theta_0 \Rightarrow \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(W, \overline{\theta})}{\partial \theta'} \xrightarrow{p} \boldsymbol{G}(\theta_0) \equiv \mathbb{E} \frac{\partial g(W, \theta_0)}{\partial \theta'}$,
7. $\boldsymbol{G}(\theta_0)$ is of full column rank.

Then

$$\sqrt{n}\left(\hat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} N\left(0, \left(\boldsymbol{G'WG}\right)^{-1} \boldsymbol{G'WSWG} \left(\boldsymbol{G'WG}\right)^{-1}\right).$$

Theorem 2: Asymptotic Distribution of Extremum Estimators

Assume that

1. $\hat{\theta} \xrightarrow{p} \theta_0$,
2. $\theta_0 \in \operatorname{int}\Theta$,
3. there exists a neighborhood $\mathcal{N}$ of θ_0 s.t. for any $\theta \in \mathcal{N}$, $Q_n(\theta)$ is twice continuously differentiable in θ for any w,
4. $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Sigma)$, Σ positive definite,
5. $\boldsymbol{H}(\theta) \equiv \frac{\partial^2 Q(\theta)}{\partial \theta \partial \theta'}$ is continuous at θ_0.
6. $\sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q(\theta)}{\partial \theta \partial \theta'} \right\| \xrightarrow{p} 0$.
7. $\boldsymbol{H}(\theta_0)$ is nonsingular.

Then

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \left(\boldsymbol{H}(\theta_0)\right)^{-1} \Sigma \left(\boldsymbol{H}(\theta_0)\right)^{-1}\right).$$

4. GMM versus ML in a classic application. Read L.P. Hansen and K.J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1285, 1982. Notice that corrections for some tables were published in *Econometrica* 52, 267-268. All texts are uploaded. If you do not know lag operators, just skim page 1279; you should be able to pick up the thread at expression 4.6.

4.1 According to Hansen and Singleton, what are major advantages of GMM over ML in their application? Explain how the authors link their argument to the "Lucas critique."

4.2 What is the intuition behind their moment conditions?

4.3 In what sense is the optimality of b_T^* "weak" (p. 1277)?

4.4 If we used GMM to estimate the fully specified model of section 4, could we match the efficiency of ML? If so, how? Does this mean that we could get the efficiency of ML without its disadvantages?

4.5 Consider the estimation results. What do α and β measure? Are the estimates plausible? How about variation between specifications?