Econ 6200: Econometrics II
Final Exam, May 15^{th}, 2024

This exam consists of 12 questions, not of equal length or difficulty, grouped into three exercises. Each question is worth 10 points. Remember to always explain your answer. Throughout, you may invoke theorems from class without proof, as well as "existence of moments as needed." Good luck!

1 (Extremum Estimators) This is an abstract question about extremum estimators defined by *minimization* of a criterion; as with similar toy examples from lecture, we abstract from randomness. Let $\Theta = [-10, 10]$ and let the sample criterion function be

$$Q_n(\theta) = |\theta^3| + n^{-1/2}(\theta - 1)^2.$$

1.1 Write down the corresponding population criterion function $Q(\theta)$ as well as the true parameter value θ_0.

1.2 Invoke a theorem from class to prove that $\hat{\theta} \equiv \arg\min_{\vartheta} Q_n(\vartheta) \xrightarrow{p} \theta_0$.

1.3 Compute $\hat{\theta}$. Use this to confirm 1.2 but also provide the order of convergence, i.e. give a statement of form "$\hat{\theta} - \theta_0 = O(n^{\alpha})$" for some α.

Reminder: The equation $ax^2 + bx + c = 0$ is solved by $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

1.4 Your result should imply that $\sqrt{n}$-consistency of $\hat{\theta}$ fails, specifically that $\sqrt{n}(\hat{\theta} - \theta_0)$ diverges. The nonrandomness is not the culprit here – I could have reverse engineered some noise so as to make $\hat{\theta}$ exactly normally distributed. What, then, goes wrong in relation to asymptotic distribution theorems from class?

2 (“Box-Cox Model”) Consider the model

$$B(Y, \lambda) = \alpha + \beta Z + \gamma B(X, \lambda) + \varepsilon,$$

where $(\varepsilon \mid Z, X) \sim N(0, \sigma^2)$ and where the function $B(\cdot)$ is given by

$$B(y, \lambda) = \begin{cases} \left(y^\lambda - 1\right)/\lambda, & \lambda \neq 0 \\ \log y, & \lambda = 0 \end{cases} .$$

For example, if $\lambda = 0$, this model specializes to the log-linear specification $\log Y = \alpha + \beta Z + \gamma \log X + \varepsilon$. You may take for granted that B is continuous in λ. Sample data are i.i.d.

2.1 Which well-known other model does this specialize to if $\lambda = 1$?

2.2 You want to test the hypothesis that the d.g.p. is described by the specialization from 2.1, while maintaining that it is accurately described by the more general model. How would you do that? Propose a test statistic that only depends on values of

$$\mathcal{L}(\lambda) = \max_{\alpha,\beta,\gamma,\sigma} \lambda \sum_{i=1}^{n} \log f(Y, Z, X; \alpha, \beta, \gamma, \lambda, \sigma^2),$$

the concentrated (with respect to all other parameters) log likelihood of the above model as a function of λ. State the statistic's precise asymptotic distribution.

2.3 For sake of argument, assume that in your data, the function $\mathcal{L}(\lambda)$ turns out to be perfectly described by $\mathcal{L}(\lambda) = -38 - 6\lambda^2 + 18\lambda$. What is the maximum likelihood estimator of λ?

2.4 Test the hypothesis from 2.2 at at 10% significance and at 5% significance. Do the answers agree?

You may assume that some important values of the standard normal c.d.f. are as follows: $\Phi(1.28) = 90\%$, $\Phi(1.64) = 95\%$, $\Phi(1.96) = 97.5\%$.

3 (Bootstrap) Consider i.i.d. observations of $Y \sim N(\mu, \sigma^2)$. You may take for granted that the simple nonparametric bootstrap analog of the sample mean converges to the appropriate normal distribution almost surely; that is, the statement

$$\sqrt{n}(\overline{Y}^* - \overline{Y}) \xrightarrow{d} N(0, \sigma^2)$$

holds with probability 1 over possible sample sequences.

Throughout this question, assume that the true expected value of Y equals $\mu_0 = 0$, but also that the researcher does not know this.

3.1 Recalling that $\mu_0 = 0$, what is the exact distribution of $n\left((\overline{Y})^2 - \mu_0^2\right)$?

(If you cannot answer this, assume henceforth that $n\left((\overline{Y})^2 - \mu_0^2\right)$ converges to a nondegenerate random variable; in particular, it neither diverges nor vanishes.)

3.2 Let

$$J_n(t) = \Pr\left(n\big((\overline{Y})^2 - \mu_0^2\big) \leq t\right)$$

be the c.d.f. of the true rescaled estimation error and let

$$J_n^*(t) = \Pr\left(n\big((\overline{Y}^*)^2 - (\overline{Y})^2\big) \leq t\right)$$

be its bootstrap analog.

Still recalling that $\mu_0 = 0$, show:

$$\begin{aligned} \lim_{n\to\infty} J_n(0) &= 0 \\ \limsup_{n\to\infty} J_n^*(0) &> 0. \end{aligned}$$

(It is ok to give a reasoned argument for the second claim rather than a fully rigorous proof.)

3.3 What does this imply for applicability of the simple nonparametric bootstrap in this example?

(If you could not show the preceding claims, assume they are true.)

3.4 Can you name one or more reasons why the result we just developed could have been guessed from information provided up to and including 3.1?

Answer Key

1.1 The term mutliplied by $n^{-1/2}$ vanishes and so the limit is $Q(\theta) = |\theta^3|$ minimized at $\theta_0 = 0$.

1.2 Contrary to how I had planned the question, both "compact domain - uniform convergence" and "convex objective - pointwise convergence" work. (Only the first one was planned. I had overlooked that $|\theta^3|$ is convex.)

1.3 $Q)n$ is convex. Also, by inspection, $\hat{\theta}$ cannot be negative because $Q_n(\theta) > Q_n(-\theta)$ for any $\theta > 0$. So we restrict attention to nonnegative arguments, meaning $|\theta^3| = \theta^3$, and evaluate the FOC

$$3\hat{\theta}^2 + 2n^{-1/2}(\hat{\theta} - 1) = 0$$
$$\implies \hat{\theta} = \frac{-2n^{-1/2} + \sqrt{4n^{-1} + 24n^{-1/2}}}{6} = \frac{-n^{-1/2} + \sqrt{n^{-1} + 6n^{-1/2}}}{3},$$

where we pick the positive root for reasons given above.

To figure out the order of this, knock out dominated terms in sums and simplify. I here spell it out in detail, though with a trained eye you'd just see it:

$$-n^{-1/2} + \sqrt{n^{-1} + 6n^{-1/2}} \approx -n^{-1/2} + \sqrt{6n^{-1/2}} \approx -n^{-1/2} + \sqrt{6}n^{-1/4} = O(n^{-1/4}).$$

You can also simplify the *abc*-formula expression to $\frac{-1+\sqrt{1+6\sqrt{n}}}{3\sqrt{n}} \approx \frac{\sqrt{6\sqrt{n}}}{3\sqrt{n}}$, which makes it arguably easier to spot.

1.4 Two traps: I pointed out that randomness is not the issue, so CLT-failure does not count and also does not tightly relate to rate of convergence. Also, the absolute value doesn't do anything here – we know the solution is in the obviously differentiable part, and on close inspection everything is differentiable anyway. (I had added the negative part of the domain to steer you away from parameter-on-boundary, though this may not have been productive.) The correct answer is that the Hessian at $\theta_0 = 0$ equals 0. In this context, check again the slide on root-n consistency without normality! That theorem would fail here too, and the slide is precise enough to tell you why.

2.1 At $\lambda = 1$, we have specialization to

$$\begin{aligned} Y - 1 &= \alpha + \beta Z + \gamma(X - 1) + \varepsilon \\ \iff \quad Y &= \underbrace{\alpha - \gamma + 1}_{\text{new intercept}} + \beta Z + \gamma X + \varepsilon, \end{aligned}$$

i.e. the model reduces to linear regression.

2.2

$$LR = 2(\max_{\lambda} \mathcal{L}(\lambda) - \mathcal{L}(1)) \xrightarrow{d} \chi_1^2.$$

2.3 This function is concave, so we take the FOC:

$$-12\hat{\lambda} + 18 = 0 \implies \hat{\lambda} = 3/2.$$

2.4 That is just executing a χ^2-test, though there were some traps. We compute

$$\begin{aligned}
&\max_{\lambda} \mathcal{L}(\lambda) = \mathcal{L}(\lambda^*) = -38 - 6 \cdot 9/4 + 27 = -11 - 54/4 = -24.5 \\
&\lambda(1) = -38 - 6 + 18 = -26 \\
\implies \quad &LR = 3
\end{aligned}$$

(first trap: did you double the difference?). The next trap is that critical values of χ_1^2 are squares of corresponding *two-sided* critical values of $N(0,1)$. Thus, we compare to $1.64^2 \approx 2.69$ versus $1.96^2 \approx 3.84$ and reject at 10% but not at 5%.

3.1 Recall that Y is exactly normal, so we have that the exact distribution $\overline{Y} \sim N(\mu_0, \sigma^2/n) \Leftrightarrow \sqrt{n}\overline{Y} \sim N(\sqrt{n}\mu_0, \sigma^2)$. But with $\mu_0 = 0$, we can then write

$$n\left((\overline{Y})^2 - \mu_0^2\right) = n(\overline{Y})^2 = \left(\sqrt{n}\overline{Y}\right)^2 \sim \sigma^2 \cdot \chi_1^2,$$

a scaled χ^2-distribution. Some important observations here are superconsistency and nonnormality. On top of that, the parameter of interest here (i.e. what we do inference on, not necessarily the notationally primitive parameter!) is $\theta = \mu^2$, and this parameter is on the boundary at 0.

3.2 We have

$$J_n(0) = \Pr(\sigma^2 \cdot chi_1^2 \leq 0) = 0,$$

so the claim holds exactly (and a fortiori in the sense of any convergence notion). Next, slightly rewrite

$$J_n^*(t) = \Pr\bigl(n(\overline{Y}^*)^2 - n(\overline{Y})^2 \leq t\bigr).$$

The intuition that would have sufficed is that we here have the difference between two r.v.'s that are not obviously ordered and so the probability of negative values should not vanish. (I agree that the formal argument to follow is a bit much for an exam, which is why I asked for an intuition, which was also spotted by a good number of students.) Somewhat more formally, from information given, we know that we can approximate

$$\begin{aligned}
\sqrt{n}\overline{Y}^* \quad &\approx \quad \sqrt{n}\overline{Y} + Z, \qquad Z \sim N(0, \sigma^2) \\
\implies n(\overline{Y}^*)^2 \quad &\approx \quad n(\overline{Y})^2 + Z\bigl(Z + 2\sqrt{n}\overline{Y}\bigr),
\end{aligned}$$

and by inspection, the last product above is negative with some positive probability; e.g., conditionally on $2\sqrt{n}\overline{Y} > 1$, it is negative whenever $Z \in (-1, 0)$, and clearly none of these events vanish.

3.3 This implies that *not* $J_n^*(t) \to J_n(t)$ in any interesting sense for $t \approx 0$ and therefore that the simple nonparametric bootstrap fails.

3.4 The red flags were listed in 3.1.