Econ 6200: Econometrics II
Final Exam, May 18^{th}, 2021

This exam consists of 12 questions, not of equal length or difficulty, grouped into four exercises. The exercises are only partly cumulative. Each question is worth 10 points. Remember to always explain your answer.

Good luck!

1 Consider the model in which

$$\begin{aligned} y &= \begin{cases} -1 & \text{if } \beta_0 + \beta_1 x - \varepsilon < -1 \\ 0 & \text{if } -1 \leq \ldots \leq 1 \\ 1 & \text{if } \beta_0 + \beta_1 x - \varepsilon > 1 \end{cases}, \\ \varepsilon &\sim N(\mu, \sigma^2). \end{aligned}$$

1.1 Impose just enough standardizations to render the model identifiable. Explain.

1.2 Propose an estimator, argue that it is consistent, and characterize its asymptotic distribution.

2 Consider a standard GMM setting. For sake of concreteness, suppose $K = 7$ and $L = 5$. Nonstandardly, suppose that the K'th moment condition is the exact unweighted average of the first $K - 1$ ones. Let the $(K - 1) \times (K - 1)$ matrix $\tilde{\boldsymbol{S}}$ denote the variance-covariance matrix of the first $(K - 1)$ moment conditions.

2.1 A researcher attempts to implement GMM ignoring the nonstandard features described above. Argue that the resulting estimate will be well-defined and consistent.

2.4 I furthermore claim that in this setting, GMM efficiency in the sense of achieving the best possible asymptotic variance is attainable even if $\hat{\boldsymbol{W}}$ is restricted to have a last row and last column of 0. Explain why this is true.

2.3 These findings are seemingly at tension with Theorems 3.1 and 3.2 because assumptions from those theorems are violated. Explain. Also explain why the theorems are nonetheless not contradicted.

3 Consider the model

$$\begin{aligned} y^* &= \beta_0 + \beta_1 x + \varepsilon \\ y &= y^* \cdot \mathbf{1}\{y^* \geq 0\}. \end{aligned}$$

Assume that (y, x) are observed, that all is i.i.d., that moments exist as needed, and also that ε is independent of x. We are only interested in the slope parameter β_1.

3.1 A researcher attempts to estimate this model by running an OLS regression of y on x. Characterize the estimator's bias as precisely as possible. (At this point, assume the distribution of x is rich enough for the asymptotic distribution to be well-defined.)

3.2 Suppose now that ε has unique median at 0. Provide conditions on the distribution of x under which the model is identified.

3.3 Illustrate your answer by constructing an instance of this model where your identifying assumption fails and where β_1 (whose value you may fix) is not identified but would be identified if y^* were observable.

3.4 Assuming your conditions hold, propose a consistent estimator and argue its consistency.
(Harder. The estimator does not have to meet an efficiency criterion. Anything that's consistent, with a plausible informal argument for its consistency, counts. You may assume that the distribution of ε is "nice" but please avoid specifying its distribution.)

4 Let the r.v. y be i.i.d. with finite moments (μ, σ^2). Consider estimation of μ by $\hat{\mu} = \overline{y}$ from an i.i.d. sample.

4.1 Prove: The idealized bootstrap standard deviation (i.e., corresponding to a bootstrap with infinite simulation size) of $\hat{\mu}$ equals $\hat{\sigma}^2/n$, where $\hat{\sigma}^2$ is a natural estimator of σ^2.

4.2 Does the estimator $\hat{\sigma}^2$ from the previous question coincide with the Maximum Likelihood estimator?

4.3 Does it coincide with the estimator that would be used in order to conduct an exact t-test (assuming this were justified)?

Quick Answers

1. Ordered Probit.

1.1 This needs location normalization, say $\mu = 0$,because there is a constant. It does *not* need a scale normalization. (Maybe it is more instructive to say I already put in the scale normalization by setting the thresholds to ± 1.)

1.2 MLE.

2.1 The estimator will be numerically equivalent to some estimator that doesn't even use the last equation because, if the last equation is a weighted average of the previous ones, I can incorporate it by accordingly changing the weights on the first $K - 1$ ones. And that's still overidentified so results from lecture apply.

2.2 Pretty much same answer. For every full-rank weighting matrix, there is one that ignores the last equation and gets the numerically same estimator.

2.3 What I was going for is simply that some relevant matrices are now positive semidefinite but not positive definite. This is not really at tension with the theorem simply because the theorem's assumptions were never claimed to be necessary conditions. Any answer which shows general understanding of what's going on here was accepted.

3.1 Write

$$y = \beta_0 + \beta_1 x + \max\{\varepsilon, -\beta_0 - \beta_1 x\},$$

and it is easy to intuit that we have attenuation bias, i.e. $\hat{\beta}_1$ is biased toward zero. It's a bit harder than intended because of the nonparametricness: that the correlation of x and $\max\{\varepsilon, -\beta_0 - \beta_1 x\}$ can be signed is intuitively compelling but slightly tricky to prove. I ignored that in grading.

3.2 We need enough distinct support points so that $\text{med}(y|x) > 0$ for two distinct values of x. That then defines two linear equations in two unknowns.

3.3 If F_x and (β_0, β_1) are such that $\text{med}(y|x) < 0$ a.s.

3.4 The assumptions allow for estimation by maximum score. So that is the easiest answer. The estimator will not be efficient though; the assumption that the distribution of ε does not depend on x can be leveraged. You had some pretty good ideas there.

4.1 The idealized bootstrap distribution is just the empirical distribution and its variance is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$, thew MLE estimator. This implies that the estimator's bootstrap variance is $\hat{\sigma}^2/n$ and the standard deviation the square root of that (you noticed the typo in the question).

4.2 Yes.

4.3 No, because that estimator has the d.f. adjustment, i.e., it equals $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$.