539B: HOMEWORK 1

Due: Wed Mar 23, 9am, via Canvas

1. We are interested in the parameters $\theta_1, \ldots, \theta_k$. θ_i may correspond, for instance, to the some measure of quality of hospital i, value added of teacher i, quality of PhD program i etc. One way to obtain estimates of these quantities is to run a panel regression in which the θ_i's would correspond to fixed effects. Based on such a regression, we obtain fixed effect estimates $\hat{\theta}_i$ that, under appropriate conditions, are asymptotically normal. For simplicity, let us assume here that the normality is exact, and that the standard errors are the same across i, let's normalize those to 1. So our model is

$$\hat{\theta}_i \sim \mathcal{N}(\theta_i, 1), \qquad i = 1, \ldots, k.$$

Often we may be interested in the rank of entity i, defined as

$$R_i = 1/2 + \sum_{j=1}^{k} \mathbb{1}\{\theta_j < \theta_i\} + \frac{1}{2} \sum_{j=1}^{k} \mathbb{1}\{\theta_j = \theta_i\}.$$

We can estimate these ranks using sample analogs,

$$\hat{R}_i = 1 + \sum_{j \neq i} \mathbb{1}\{\hat{\theta}_j < \hat{\theta}_i\} + \frac{1}{2} \sum_{j \neq i} \mathbb{1}\{\hat{\theta}_j = \hat{\theta}_i\}. \tag{1}$$

How should we construct confidence intervals for $\hat{R}_i$? One idea is to report percentile bootstrap confidence intervals (CIs), or Efron's percentile bootstrap CIs based on the parametric bootstrap, where the bootstrap draws are given by $\hat{\theta}_i^* \sim \mathcal{N}(\hat{\theta}_i, 1)$, $i = 1, \ldots, k$.

   (a) Are either or both versions (the percentile method or Efron's percentile method) of the bootstrap CI going yield CIs with asymptotically correct coverage? (here $k = k_n$ could go to ∞ as the sample size $n \to \infty$, or else $\operatorname{var}_n(\hat{\theta}_i) \to 0$, or both). Explain briefly your reasoning (no math needed). If your answer is yes, are there some data generating processs (DGPs) that you need to rule out? If your answer is no, explain why—be specific, explain what goes wrong if possible (it's not enough to say, for instance, that you don't think the distribution of $\hat{R}_i$ will be asymptotically normal).

(b) Give an exact algorithm for how you'd construct such bootstrap CIs. Be clear enough so that your undergraduate research assistant would understand what to do.

(c) Unfortunately, you do not have such an assistant. So to verify your answer to 1a, you'll need to do some simulations yourself. Let us consider $k = 10$, and three designs: in the first design, let $\theta_i = i$. In the second design, $\theta_i = 10i$, and in the third design, $\theta_i = i/10$. Consider $M = 5,000$ simulation draws for each design. For each simulation draw m, construct a percentile and Efron's CIs $\hat{CI}^{perce}$ and $\hat{CI}^{efron}_{m,i}$ for the ranks R_i, $i = 1, \ldots, k$. Use $N = 1,000$ bootstrap draws per simulation draw, and nominal 95% level.

Report coverage of both methods across the designs (your table should have 60 numbers between 0 and 1). Update your answer to 1a as needed.

(d) Do you notice anything undesirable about the form of the percentile bootstrap CI in this case?

2. Bergeron, Tourek, and Weigel (2020) analyze an randomized controlled trial (RCT) in which the property tax faced by different households was randomized in the city of Kananga. They are interested in estimating the elasticity of tax compliance. They run the following regression:

$$Y_i = \alpha + \beta X_i + \sum_h \mathbb{1}\{h(i) = h\}\gamma_h + \epsilon_i,$$

where Y_i is an indicator that individual i paid the property tax, X_i is the (randomized) property tax that the individual faced, $h(i)$ is the neighborhood of individual i (so γ_h are neighborhood effects), and ϵ_i is regression error. They are interested in the elasticity of tax compliance, which corresponds to the parameter $\theta = \beta/E[Y_i]$. They estimate this elasticity as $\hat{\theta} = \hat{\beta}/\bar{Y}$, where $\hat{\beta}$ is an ordinary least squares (OLS) estimate. They use the (nonparametric) bootstrap to estimate the standard error of $\hat{\theta}$.

(a) Does this approach yield (asymptotically) correct standard errors? If your answer is yes, are there some DGPs that you need to rule out? If your answer is no, explain why—be specific, explain what goes wrong if possible (it's not enough to say, for instance, that you don't think the distribution of $\hat{\theta}$ will be asymptotically normal).

(b) Suppose you don't have the computing power to implement the bootstrap. How else could you compute the standard error of $\hat{\theta}$? Explain. If your answer to (a) is that we need to rule out some DGPs, is it possible to construct CIs that are valid even under these problematic DGPs?

REFERENCES

Bergeron, Augustin, Gabriel Tourek, and Jonathan L Weigel. 2020. "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC." Working paper, Harvard University, November. https://www.dropbox.com/s/vpdtonwqqnp5nwt/Randomized_Rates_Paper_JMP.pdf.