THE BOOTSTRAP AND SUBSAMPLING

Michal Kolesár*

March 12, 2025

In this lecture, we discuss the bootstrap: a resampling method for approximating the distribution (or particular features of the distribution) of a given statistic. First we discuss the mechanics of how it works and how to implement it in practice. Second, we discuss why one may want to use it. Finally, we discuss situations in which it doesn't work at all, and what one can do in such situations.

1. THE BOOTSTRAP

1.1. *The usual approach to inference*

Let us first review the usual approach to inference, using slightly different notation to set us up for the bootstrap. I follow the setup in Lehmann and Romano (2005, Chapter 15.4). We observe $X_i \sim F \in \mathcal{F}$, $i = 1, \dots, n$. The parameter space $\mathcal{F}$ can be parametric or non-parametric. Parametric just means here that there is a finite-dimensional parameter $\gamma \in \Gamma \subseteq \mathbb{R}^k$ such that $\mathcal{F} = \{F_\gamma \colon \gamma \in \Gamma\}$; the parameter space is non-parametric otherwise. We're interested in some parameter $\theta(F) \in \Theta = \{\theta(F) \colon F \in \mathcal{F}\} \subseteq \mathbb{R}^d$. Denote the sample by $\mathbf{X}_n = (X_1, \dots, X_n)$.

We'd like to construct a confidence interval (CI) for θ using the real-valued function $R_n(\mathbf{X}_n, \theta(F))$. In the bootstrap literature, this function is called a *root* (the name is due to Beran 1984). In principle, R_n can be any real-valued function, but in practice, it typically takes one of the four following forms:

Example 1. If we have available a $\sqrt{n}$-consistent estimator $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n)$ of θ, then there are two options for the root:

1. $R_n(\mathbf{X}_n, \theta) = \sqrt{n}(\hat{\theta}_n - \theta)$; or
2. $R_n(\mathbf{X}_n, \theta) = |\sqrt{n}(\hat{\theta}_n - \theta)|$.

If we also have an estimator of its asymptotic variance, $\hat{\sigma}_n^2 = \hat{\sigma}^2(\mathbf{X}_n)$, then there are two further options:

*Email: mkolesar@princeton.edu

3. the t-statistic, $R_n(\mathbf{X}_n, \theta) = \sqrt{n}(\hat{\theta}_n - \theta)/\hat{\sigma}_n$,
4. the absolute value of the t-statistic, $R_n(\mathbf{X}_n, \theta) = \sqrt{n}|\hat{\theta}_n - \theta|/\hat{\sigma}_n$. ☒

Since both $\mathbf{X}_n$ and θ depend on F, the distribution of R_n will also in general depend on F. Let $J_n(\cdot, F)$ denote this distribution, so that $J_n(r, F) = P_F(R_n \leq r)$. Define the quantiles of J_n,

$$J_n^{-1}(\alpha, F) = \inf\{r\colon J_n(r, F) \geq \alpha\}.$$

If we knew these quantiles, we could construct a (finite-sample valid) CI for θ as

$$\mathrm{CI}^* = \{\theta \in \Theta\colon R_n(\mathbf{X}_n, \theta) \leq J_n^{-1}(1-\alpha, F)\},$$

or, depending on the root, we could make use of both tails of the distribution, yielding

$$\widetilde{\mathrm{CI}}^* = \{\theta \in \Theta\colon J_n(\alpha/2, F) \leq R_n(\mathbf{X}_n, \theta) \leq J_n^{-1}(1-\alpha/2, F)\}.$$

Example 1 (continued). For two-sided CIs, using CI^* for cases 2 and 4, and $\widetilde{\mathrm{CI}}^*$ for cases 1 and 3, we get

$$\begin{aligned}
\widetilde{\mathrm{CI}}_1^* &= [\hat{\theta}_n - J_n^{-1}(1-\alpha/2, F)/\sqrt{n}, \hat{\theta}_n - J_n(\alpha/2, F)/\sqrt{n}],\\
\mathrm{CI}_2^* &= \{\hat{\theta}_n \pm J_n^{-1}(1-\alpha, F)/\sqrt{n}\},\\
\widetilde{\mathrm{CI}}_3^* &= [\hat{\theta}_n - J_n^{-1}(1-\alpha/2, F)\hat{\sigma}_n/\sqrt{n}, \hat{\theta}_n - J_n(\alpha/2, F)\hat{\sigma}_n/\sqrt{n}],\\
\mathrm{CI}_4^* &= \{\hat{\theta}_n \pm J_n^{-1}(1-\alpha, F)\hat{\sigma}_n/\sqrt{n}\}.
\end{aligned}$$

Note the reversal of the quantiles! ☒

This observation leads to the idea of constructing CIs using the pivot method, which you saw in 517. The idea of the pivot method is to be clever about choosing the root appropriately so that its distribution J_n doesn't depend on F.

Example 2. Suppose $X_i \sim \mathcal{N}(\theta, \sigma^2)$, $i = 1, \dots, n$. If we set $R_n = \sqrt{n}(\overline{X}_n - \theta)/S_n$, where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ is the sample mean, and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ is the sample variance, then the t-statistic R_n is distributed t with $n-1$ degrees of freedom, independently of θ or σ^2. So R_n is a pivot. Similarly, $R_n = |\sqrt{n}(\overline{X}_n - \theta)/S_n|$ is a pivot. Plugging in quantiles of the t-distribution to $\widetilde{\mathrm{CI}}_3^*$ or CI_4^* then leads to the usual two-sided CI, $\overline{X}_n \pm t_{1-\alpha/2}(n-1) \cdot S_n/\sqrt{n}$. ☒

Exact pivots are rare. The standard way of doing inference (e.g. inference based on extremum estimators) is to find R_n that is asymptotically pivotal, that is find R_n such that $J_n(r, F) \to J_\infty(r)$ at all continuity points of J_∞ (this is just the definition of convergence in distribution). Then, provided J_∞ is continuous, we can use quantiles of J_∞ for inference.[1]

1. By Lemma 21.2 in van der Vaart (1998), $J_n(r, F) \to J_\infty(r)$ at all continuity points of J_∞ if and only if $J_n^{-1}(r, F) \to J_\infty^{-1}(r)$ at all continuity points.

Example 1 (continued). In "regular" models, we are typically able to show that $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \sigma^2)$, and that $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$, so that for all $F \in \mathcal{F}$, $R_n = \sqrt{n}(\hat{\theta}_n - \theta)/\hat{\sigma}_n \Rightarrow \mathcal{N}(0, 1)$, and J_∞ is thus the standard normal distribution. This leads to the "standard" asymptotic CIs for θ given by

$$\widetilde{\mathrm{CI}}_3^A = \mathrm{CI}_4^A = \{\hat{\theta}_n \pm z_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}\}. \qquad \boxtimes$$

1.2. *The bootstrap idea*

The bootstrap idea is due to Efron (1979), who gave it the name.[2] The idea of the bootstrap is to approximate the distribution $J_n(\cdot, F)$ by $\hat{J}_n = J_n(\cdot, \hat{F}_n)$, where $\hat{F}_n$ is some estimate of F. There are many versions of the bootstrap, each with different estimator of F. In the *nonparametric (or empirical) bootstrap* considered in Efron (1979), $\hat{F}_n$ is the empirical cumulative distribution function (CDF). The *parametric bootstrap* uses the estimator $F_{\hat{\gamma}_n}$, where $\hat{\gamma}_n$ is some reasonable estimator of γ.

Question 1. Such as?

This yields the bootstrap CIs

$$\mathrm{CI}^B = \{\theta \in \Theta \colon R_n(\mathbf{X}_n, \theta) \leq \hat{J}_n^{-1}(1-\alpha)\},$$

or, using both tails of the distribution,

$$\widetilde{\mathrm{CI}}^B = \{\theta \in \Theta \colon \hat{J}_n(\alpha/2) \leq R_n(\mathbf{X}_n, \theta) \leq \hat{J}_n^{-1}(1-\alpha/2)\}.$$

In practice, it is hard to compute the quantiles $\hat{J}_n^{-1}(\cdot)$ directly (for the nonparametric bootstrap, a brute force approach requires evaluation of n^n samples), but one can always use simulation. In particular, we can:

1. Draw a sample $\mathbf{X}_n^*$ of size n from $\hat{F}_n$. For the nonparametric bootstrap, this just means sampling from the observed data $\mathbf{X}_n$ with replacement. For the parametric bootstrap, if we have a parametric model $F_\gamma(\cdot \mid z)$ for an outcome Y given a vector of regressors Z, this amounts to drawing $\mathbf{Z}_n^*$ with replacement from the empirical distribution, and drawing the outcomes Y_i^* from the distribution $F_{\hat{\gamma}}(\cdot \mid Z_i^*)$. Alternatively, one can put $\mathbf{Z}_n^* = \mathbf{Z}_n$

   Question 2. What's the difference?

2. Compute $R_n^* = R_n(\mathbf{X}_n^*, \theta(\hat{F}_n))$.

2. Tukey suggested "shotgun" because, as explained in Efron (1979), the bootstrap can "blow the head off any problem if the statistician can stand the resulting mess".

3. Repeat N times to obtain $(R^*_{n,1}, \ldots, R^*_{n,N})$. Use quantiles of this empirical distribution to approximate $\hat{J}_n^{-1}(t)$. We can make the approximation arbitrarily close by making N large.

Example 2 (continued). Suppose we didn't realize that $\sqrt{n}(\overline{X}_n - \theta)/S_n$ was a pivot, and instead we bootstrapped $R_n = \sqrt{n}(\overline{X}_n - \theta)$. Consider first the nonparametric bootstrap. Under the empirical CDF $\hat{F}_n$, the mean is $\theta(\hat{F}_n) = \int x \, \mathrm{d}\hat{F}_n(x) = \overline{X}_n$, so that $R^*_n = \sqrt{n}(\overline{X}^*_n - \overline{X}_n)$, where $\overline{X}^*_n = n^{-1}\sum_i X^*_i$ and X^*_i are drawn from $\mathbf{X}_n$ with replacement. This yields (note the reversal of the "natural" quantiles)

$$\widetilde{\mathrm{CI}}^B = [\overline{X}_n - \hat{J}_n^{-1}(0.975)/\sqrt{n}, \overline{X}_n - \hat{J}_n^{-1}(0.025)/\sqrt{n}].$$

Now consider the parametric bootstrap. Suppose we use the estimators $\hat{\gamma}_n = (\overline{X}_n, S_n^2)$ of θ and σ^2. Again, $\theta(F_{\hat{\gamma}_n}) = \overline{X}_n$. Thus, $R^*_n = \sqrt{n}(\overline{X}^*_n - \overline{X}_n)$, where X^*_i is drawn from $\mathcal{N}(\overline{X}_n, S_n^2)$. Thus, $R^*_n \sim \mathcal{N}(0, S_n^2)$. Therefore, the bootstrap CI is given by $\overline{X}_n \pm z_{1-\alpha/2} S_n/\sqrt{n}$, which is just the usual CI (except we use quantiles of the normal, rather than those of the t-distribution). ⊠

Example 1 (continued). Unless $\hat{\theta}_n$ is a plug-in estimator, $\hat{\theta}_n = \theta(\hat{F}_n)$, it will not generally be the case that $\theta(\hat{F}_n) = \hat{\theta}_n$. Nonetheless, to construct CIs for θ_n, people often approximate the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ by the bootstrap distribution of $R^*_n = \sqrt{n}(\hat{\theta}_n(\mathbf{X}^*_n) - \hat{\theta}_n)$. This leads to the *percentile bootstrap CI* (note again the reversal of the quantiles)

$$\widetilde{\mathrm{CI}}_1^B = [\hat{\theta}_n - \hat{J}_n^{-1}(1-\alpha/2)/\sqrt{n}, \hat{\theta}_n - \hat{J}_n^{-1}(\alpha/2)/\sqrt{n}],$$

a bootstrap analog of $\widetilde{\mathrm{CI}}_1^*$. Using the t-statistic instead leads to the *percentile-t bootstrap CI*. In this case, we take bootstrap draws $R^*_n = \sqrt{n}(\hat{\theta}_n(\mathbf{X}^*_n) - \hat{\theta}_n)/\hat{\sigma}_n(\mathbf{X}^*)$, leading to the CI

$$\widetilde{\mathrm{CI}}_3^B = [\hat{\theta}_n - \hat{J}_n^{-1}(1-\alpha/2)\hat{\sigma}_n/\sqrt{n}, \hat{\theta}_n - \hat{J}_n^{-1}(\alpha/2)\hat{\sigma}_n/\sqrt{n}],$$

which is similar to the usual CI ($\widetilde{CI}_3^A$), except we replaced the normal quantiles $z_{1-\alpha/2}$ and $z_{\alpha/2}$ (i.e. 1.96 and -1.96 for $\alpha = 0.05$) with quantiles from the bootstrap distribution. In practice, people often instead use the CI $[\hat{\theta}_n + \hat{J}_n^{-1}(\alpha/2)/\sqrt{n}, \hat{\theta}_n + \hat{J}_n^{-1}(1-\alpha/2)/\sqrt{n}]$, which is known as *Efron's percentile method*. The logical justification for this method is less strong, but it often works fine in practice, and the difference is asymptotically negligible if the asymptotic distribution of $\hat{\theta}_n - \theta$ is symmetric around zero. One could also of course bootstrap the absolute values, leading to the CIs:

$$\mathrm{CI}_2^B = \{\hat{\theta}_n \pm \hat{J}_n^{-1}(1-\alpha)/\sqrt{n}\},$$
$$\mathrm{CI}_4^B = \{\hat{\theta}_n \pm \hat{J}_n^{-1}(1-\alpha)\hat{\sigma}_n/\sqrt{n}\},$$

which ensures that the CI is symmetric around $\hat{\theta}_n$. ⊠

In practice, the most common method of using the bootstrap is to use it to estimate the standard error of an estimator, so that the reader can form a CI at whatever level they want. For this to work, of course, we need that $R_n = \sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically normal. Then:

1. Draw a sample $\mathbf{X}_n^*$ of size n from $\hat{F}_n$, and compute $\hat{\theta}_n(\mathbf{X}_n^*)$.

2. Do this N times to obtain $(\hat{\theta}_n(\mathbf{X}_{n,1}^*), \ldots, \hat{\theta}_n(\mathbf{X}_{n,N}^*))$. Use the sample standard deviation $\hat{\sigma}_{n,*} = (N^{-1}\sum_{j=1}^N(\hat{\theta}_n(\mathbf{X}_{n,j}^*) - \overline{\hat{\theta}_n}^*)^2)^{1/2}$, $\overline{\hat{\theta}_n}^* = N^{-1}\sum_{j=1}^N \hat{\theta}_n(\mathbf{X}_{n,j}^*)$ of this bootstrap sample, an estimate of the standard error of $\hat{\theta}_n$, which leads to the CI

$$\mathrm{CI}_5^B = \{\hat{\theta}_n \pm z_{1-\alpha/2}\hat{\sigma}_{n,*}\}.$$

A variant of this approach is to approximate the second moment of R_n, replacing $\hat{\sigma}_{n,*}^2$ with $\hat{s}_{n,*}^2 = \frac{1}{N}\sum_{j=1}^N(\hat{\theta}_n(\mathbf{X}_{n,j}^*) - \hat{\theta}_n)^2$

Question 3. What if $\hat{\theta}_n$ is not asymptotically normal? What is the relationship between $\hat{\sigma}_{*,n}$ and $\hat{s}_{n,*}$?

Example 2 (continued). We want to bootstrap the standard error of $\overline{X}_n$. We know that for the non-parametric bootstrap, the bootstrap distribution has mean $\overline{X}_n$, and second moment $n^{-1}\sum_i X_i^2$ (why?). So $\hat{\sigma}_{n,*}^2 = \mathrm{var}(\overline{X}_n^*) = \mathrm{var}(X_i^*)/n = \sum_i(X_i - \overline{X}_n)^2/n^2$, which is the usual standard error for $\overline{X}_n$. What if we use the parametric bootstrap? ⊠

1.3. *Consistency of the bootstrap*

It's usually the case that the R_n converges in distribution to some limit, so that $J_n(t, F) \to J_\infty(t, F)$ at all continuity points t of J_∞. Also, $\hat{F}_n$ converges to F for most reasonable estimators $\hat{F}_n$. For instance, for the nonparametric bootstrap, we know that $\|\hat{F}_n - F\|_\infty \overset{\text{a.s.}}{\to} 0$ by the Glivenko-Cantelli theorem, where $\|F\|_\infty = \sup_x|F(x)|$ is the supremum norm. Thus, $\hat{J}_n = J_n(\cdot, \hat{F}_n)$ should be close to $J_\infty(\cdot, F)$, provided that $J_n(t, F)$ is suitably continuous in its second argument. So proving that the bootstrap distribution will be close to J_∞, the asymptotic distribution of R_n requires three ingredients:

1. $\hat{F}_n$ converges to F

2. $J_n(\cdot, F)$ converges to $J_\infty(\cdot, F)$ (i.e. there exists a well-defined asymptotic distribution).

3. J_n is "continuous" in the sense that if F_n converges to F, then $J_n(\cdot, F_n)$ converges to $J_\infty(\cdot, F)$.

If J_∞ is continuous, the consistency of the bootstrap, that is

$$J_n(x, \hat{F}_n) \overset{\text{a.s.}}{\to} J_\infty(x, F) \quad \text{for all } x \tag{1}$$

then follows by an extended version of a continuous mapping theorem.[3] It is then not much more work to show that this implies that this implies consistency of the bootstrap CIs $\widetilde{\mathrm{CI}}_1^B, \mathrm{CI}_2^B, \widetilde{\mathrm{CI}}_3^B, \mathrm{CI}_4^B$. See for example, Politis, Romano, and Wolf (1999, Theorem 1.2.1).

Remark 1 (Bootstrap standard error). For showing consistency of bootstrapped standard errors (i.e. that $n\hat{\sigma}_{n,*}^2 \xrightarrow{p} \sigma^2$, if $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \sigma^2)$), we need slightly stronger conditions than showing that the percentile bootstrap leads to asymptotically valid CIs. This is because convergence in distribution doesn't imply convergence of moments (we need a uniform integrability condition in addition, which may be difficult to establish). In fact, by arguments you may have seen in 517 (Remark 6 of 517 lecture note 3 that's on GitHub), the asymptotic variance is not necessarily the limit of finite-sample variances, instead $\liminf_{n\to\infty} \operatorname{var}(\sqrt{n}(\hat{\theta}_n - \theta)) \geq \sigma^2$. Hahn and Liao (2021) use the same observation to show that CI_5^B is generally conservative—the bootstrap standard error is too large in general. Much like the Monte Carlo standard error for an instrumental variables (IV) estimator will be too large in general relative to its asymptotics variance, the bootstrap variance in IV will be too large in general. If you want to report a standard error, a better idea is to take one of the bootstrap CIs and divide its length by $2 \times z_{1-\alpha}$ (take, say $\alpha = 0.05$, or put $\alpha = 0.25$, which corresponds to using interquartile range to back out the asymptotic variance).

Remark 2 (Rule of thumb). If $R_n(\mathbf{X}_n, \theta) = \sqrt{n}(\hat{\theta}_n - \theta)$, and $\hat{\theta}_n$ is an extremum estimator that's asymptotically linear (what does that mean?) by the large-sample distribution theorems for extremum estimators that you may have discussed in 518 or 519, then the bootstrap is typically valid.[4] For example, the fact that the percentile bootstrap for quantile regression is valid was shown in Hahn (1995) (this was part of his PhD dissertation). Conversely, if the distribution is not asymptotically linear, the bootstrap will typically fail. For example, the bootstrap fails to consistently estimate the asymptotic distribution of the maximum score estimator, as first pointed out in Abrevaya and Huang (2005).[5]

Research question: Is it possible to formalize some part of this rule of thumb?

Remark 3 (Parametric vs. nonparametric bootstrap). The parametric bootstrap exploits the parametric model structure, and so will often perform better than the nonparametric bootstrap. On the other hand, if the model used in the parametric bootstrap is misspecified, then the parametric bootstrap will generally be inconsistent, while the nonparametric bootstrap will typically still yield correct inference for the appropriate pseudo-parameter.

3. The theorem says (see van der Vaart 1998, Theorem 18.11) that if $g_n(x_n) \to g(x)$ whenever $x_n \to x$, then $X_n \overset{\text{a.s.}}{\to} X$ implies $g_n(X_n) \overset{\text{a.s.}}{\to} g(X)$. Here X_n are random elements.

4. This rule of thumb is based on Theorem 1 in Mammen (1992, p. 9): Consider bootstrapping a linear functional of F (i.e. for some h, $\theta = \int h(x)\,\mathrm{d}F(x)$, and $\hat{\theta}_n = n^{-1}\sum_i h(x_i)$). Then the non-parametric bootstrap is consistent if and only if $\hat{\theta}_n$ is asymptotically normal. This suggests the bootstrap should be consistent if the statistic of interest is linear asymptotically, $\sqrt{n}(\hat{\theta}_n - \theta) = n^{-1/2}\sum_i h(x_i; \theta) + o_p(1)$, as discussed in 519.

5. It appears that some of their results are incorrect; see Kosorok (2008) and Sen, Banerjee, and Woodroofe (2010) for discussion as well as cleaner arguments for bootstrap inconsistency in a related class of models. See also Cattaneo, Jansson, and Nagasawa (2020) for a modification of the bootstrap to restore its validity.

2. WHY USE THE BOOTSTRAP

So there are three different bootstrap CIs: the percentile bootstrap CI $\widetilde{\mathrm{CI}}_1^B$ (or the symmetric version CI_2^B), the percentile-t bootstrap $\widetilde{\mathrm{CI}}_3^B$ (or the symmetric version CI_4^B), and CI_5^B, based on bootstrapping the standard errors. If the asymptotic distribution is normal, we can also use the usual CI, $\widetilde{\mathrm{CI}}_3^A$. We now discuss the relative merits of the different CIs.

2.1. *Why use the percentile bootstrap*

A clear advantage of the $\widetilde{\mathrm{CI}}_1^B$ is that it doesn't require the user to construct a consistent estimator of the asymptotic variance. This is especially useful when an analytic approach to constructing such an estimator is complicated or intractable, as the following two examples demonstrate.

Example 3 (Quantile regression). Consider the linear quantile regression model. That is, letting $Q_\tau(Y \mid X) = F_{Y|X}^{-1}(\tau \mid X)$ denote the conditional quantile function of Y given X at quantile τ, suppose $Q_\tau(Y \mid X) = X'\theta$. Recall from 517 that $Q_\tau(Y \mid X)$ solves the minimization problem $\min_{q(X)} E\rho_\tau(Y - q(X))$, where

$$q_\tau(u) = (\tau - \mathbb{1}\{u \leq 0\})u = \tau\,\mathbb{1}\{u > 0\}u + (1-\tau)\,\mathbb{1}\{u \leq 0\}(-u).$$

is the check function. The quantile regression estimator minimizes the sample analog:[6]

$$\hat{\theta}_n = \operatorname*{argmin}_\theta \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_i - X_i'\theta).$$

One can show (using the tools discussed in 519) that

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \tau(1-\tau)E[f_u(0 \mid X_i)X_iX_i']^{-1}E[X_iX_i']E[f_u(0 \mid X_i)X_iX_i']^{-1}),$$

where $f_u(0 \mid X)$ is the conditional density of the residual $u_i = Y_i - X_i'\theta$ at 0. Here the asymptotic variance is tricky to estimate; we'd like to avoid having to obtain nonparametric estimates of conditional densities. The bootstrap allows us to do exactly that (in Stata, `bsqreg` implements these bootstrapped standard errors). ⊠

Example 4 (Counterfactuals). Consider a discrete choice problem, where $U_{ij} = X_{ij}'\theta + \epsilon_{ij}$ is the utility of person i from choice $j = 0, \ldots, J$. Suppose that ϵ_{ij} has a type-I extreme value distribution $F(\epsilon) = e^{-e^{-\epsilon}}$. Person i makes choice j if and only if $U_{ij} \geq U_{ik}$ for $k = 0, \ldots, J$. This is called the conditional logit model. By evaluating some integrals,

6. In practice, we can compute $\hat{\theta}_n$ by casting the minimization problem as a linear programming problem:

$$\min_{\theta, u_+, u_-} \tau\iota' u_+ + (1-\tau)\iota' u_-, \qquad Y = X'\theta + u_+ + u_-, \qquad u_+, u_- \geq 0.$$

one can show that the *conditional choice probabilities* are given by

$$P(Y_i = j) = \frac{e^{X'_{ij}\theta}}{\sum_{k=0}^{J} e^{X'_{ik}\theta}}.$$

We can use this to build a likelihood (how?) and estimate θ. What happens if $J = 1$? The coefficients are hard to interpret, and typically not themselves of interest. Instead, we're interested in some counterfactuals, or average elasticities, $E[h(X_{i0}, \ldots, X_{iJ}, \theta)]$, or $n^{-1}\sum_i h(X_{i0}, \ldots, X_{iJ}, \theta)$, where h is some complicated function (Exercise: suppose we want to compute the average own price elasticity; how does h look like?). So even if we make use of the usual theory to get standard errors for $\hat{\theta}_n$, using the delta method may be tedious. This issue is even more severe in more complicated discrete choice models.⊠

Question 4. What is the general lesson here?

$\widetilde{\mathrm{CI}}_1^B$ has two further advantages relative to the percentile-t bootstrap and the usual CI:

1. It is robust to model misspecification. If the model is misspecified, it still appropriately reflects the sampling uncertainty, if we view $\hat{\theta}_n$ as an estimator of the pseudo-parameter. In contrast, the percentile-t bootstrap and the usual CI don't have this interpretation unless we use misspecification-robust standard errors.

2. Because implementing the bootstrap resampling is typically much simpler than implementing asymptotic variance estimators, the approach is also more robust to coding errors.

2.2. *Why use the percentile-t bootstrap*

If we bootstrap t-statistics, then it will typically be the case that $\hat{J}_n$ is consistent and J_∞ is pivotal. So one may use either one as an approximation to J_n, and it's not clear which approximation is better, although it's clear that the bootstrap critical values are more work.

In some cases, however, one can show that the bootstrap approximation is better, in which case the bootstrap is said to provide an *asymptotic refinement*. The intuition is that while the finite-sample distribution of an estimator will be skewed, the normal asymptotic approximation is symmetric. On the other hand, the bootstrap may be able to capture the skewness. Typically, a necessary condition for showing that the bootstrap gives such a refinement is that J_∞ is pivotal. Takeaway: it's better to bootstrap t-statistics, if you can, rather than estimators or standard errors.

Formalizing the intuition requires Edgeworth expansions—refinements of a central limit theorem (CLT). The classic reference for this is Hall (1992). The main takeaways are:

1. Typically, if J_∞ is pivotal, then the bootstrap error in approximating the distribution of $J_n(\cdot, F)$ is of the order $O(n^{-1})$, while the CLT approximation (or the percentile bootstrap) has error $O(n^{-1/2})$.

   More technical arguments are needed to show that this better approximation to the asymptotic distribution translates into more accurate coverage probabilities for confidence intervals.

2. To be able to do these Edgeworth expansion in the first place, we need more moment assumptions. For example, in the case of bootstrapping the mean, Hall (1988) showed that finiteness of three absolute moments is in fact necessary and sufficient for higher-order accuracy of the bootstrap (and if the third moment doesn't exist, the bootstrap can in fact do worse than the CLT).

3. Furthermore, we also need $J_n(\cdot, F)$ to be smooth enough. For example, while the bootstrap is consistent in Example 3, the bootstrap error is of bigger order than the usual rate $O(n^{-1/2})$; (de Angelis, Hall, and Young 1993). One could smooth the objective function or smooth the bootstrap to improve the rate (see Section 4 in Horowitz (2019) for a discussion), but doing so requires a judicious choice of bandwidth.

Remark 4. See section 3.7 of Horowitz (2001) for discussion of how to get a refinement in generalized method of moments (GMM) models: the usual approach does not give it. The original reference is Hall and Horowitz (1996).

2.3. *Less common reasons to use the bootstrap*

1. The bootstrap is also used to provide critical values for testing procedures. For instance, Kitagawa (2015) uses the bootstrap to develop a test for instrument validity that allows for heterogeneous treatment effects.

2. The bootstrap can also sometimes estimate the bias of an estimator up to some asymptotic order. The bootstrap bias estimate can then be subtracted from the original parameter estimator to give rise to a less biased estimator (at the expense of increasing the variance). See Horowitz (2001, Chapter 3.1).

   A very nice application of this idea is Higgins and Jochmans (2024). In parametric non-linear panel data models with fixed effects (e.g. logit), if $T \to \infty$ maximum likelihood estimators of the common parameters θ will be asymptotically normal. However, unless T grows faster than n, the estimator will be asymptotically biased: $\sqrt{nT}(\hat{\theta} - \theta) \approx_a \mathcal{N}(\sqrt{n/T}b, \Sigma)$. This bias b arises due to the fixed effects nuisance parameter, the dimension of which grows with n. One could devote a lot of effort to figuring out the form of the bias, and bias-correct the estimator; but this is tedious and one has to re-do it for each model.

Higgins and Jochmans (2024) observe that this effort can be avoided, because the parametric bootstrap will actually replicate this bias, so that $\sqrt{nT}(\hat{\theta}^* - \theta) \approx_a \mathcal{N}(\sqrt{n/T}b, \Sigma)$. Therefore, the percentile bootstrap CI $\widetilde{\text{CI}}_1^B$ (or the symmetric version CI_2^B) will be asymptotically valid. In other words, the bootstrap critical values $\hat{J}_n^{-1}(1-\alpha/2)$ will be *bias-aware*: correspond to quantiles of the biased distribution $\mathcal{N}(\sqrt{n/T}b, \Sigma)$. We'll see such bias-aware critical values again when we talk about regression discontinuity (but we'll have to compute them using other methods, the bootstrap will not work there). If one is interested in estimation, one can take the median of the bootstrap distribution as a bias estimate, and subtract it off.

3. We know that in nice models, the bootstrap and the conventional normal asymptotic approximation should agree in large samples under conventional asymptotics. Andrews and Shapiro (2024) explore the idea of using the discrepancy between the (Bayes) bootstrap distribution and the asymptotic approximation as a way of gauging the quality of the normal approximation.

2.4. *Key things to remember*

- A key requirement for bootstrap validity is that the bootstrap distribution $\hat{F}_n$ is close to the true DGP F. This means that in models with dependence, the bootstrap needs to respect the dependence. For example, with clustered data, we need to bootstrap the clusters. In panel applications, we need to resample the individuals, preserving the time-series dependence of the data. In time-series applications, we resample blocks of the data ("block bootstrap") to preserve the dependence structure in the data

- You only get an asymptotic refinement with the percentile-t method, which is asymptotically pivotal. Otherwise, if the percentile method, or the bootstrapped standard errors yield a CI that's different from the usual CI, it is not clear which one we should prefer—we just know that there probably an issue with the asymptotic approximation (but the reverse is false in general!). Remember the reversal of the quantiles (bootstrap the absolute value of the t-statistics if you want the CI to be symmetric).

- Don't bootstrap statistics that are non-smooth or not asymptotically normal (or, more precisely, asymptotically linear) without due diligence.

- Use the sandwich formula when bootstrapping t-statistics with the non-parametric bootstrap.

3. FAILURE OF THE BOOTSTRAP

Provided that we respect the structure of the data as discussed in Section 2.4, it follows from the discussion in Section 1.3 that bootstrap failure occurs if, heuristically, small differences between $\hat{F}_n$ and F can translate into large differences between $J_n(t, \hat{F}_n)$ and $J_n(t, F)$. We now go through some examples of when this happens. All of these examples are in line with the "rule of thumb" in Remark 2.

Example 5 (Parameter on the boundary). This example is due to Andrews (2000). Suppose $X_i \sim F$, we're interested in the mean $\theta(F) = \int x \, \mathrm{d}F(x)$, and we know that $\theta(F) \geq 0$. Suppose $\mathrm{var}_F(X_i) = 1$.

- A natural estimator is $\hat{\theta}_n = \max(0, \overline{X}_n)$. Let $R_n = n^{1/2}(\hat{\theta}_n - \theta)$.
- If $\theta > 0$, then $R_n \Rightarrow Z \sim \mathcal{N}(0,1)$, but if $\theta = 0$, then $R_n \Rightarrow \max\{Z, 0\}$.

Now, suppose that $\theta = 0$ and fix $c > 0$. Then, on the event $A_n = \{\sqrt{n}\overline{X}_n \geq c\}$,

$$R_n^* = n^{1/2}(\hat{\theta}_n^* - \theta(\hat{F}_n)) = n^{1/2}(\max\{\overline{X}_n^*, 0\} - \overline{X}_n) = \max\{n^{1/2}(\overline{X}_n^* - \overline{X}_n), -n^{1/2}\overline{X}_n\}$$
$$\leq \max\{n^{1/2}(\overline{X}_n^* - \overline{X}_n), -c\}.$$

We know that conditional on $\mathbf{X}_n$, $n^{1/2}(\overline{X}_n^* - \overline{X}_n)$ is standard normal in large samples by CLT, so the RHS in the display above converges to $\max\{Z, -c\}$. Consequently, $\lim_{n\to\infty} P(R_n^* \leq x \mid A_n) \geq P(\max\{Z, -c\} \leq x) > P(\max\{Z, 0\} \leq x) = J_\infty(x)$. Thus, the bootstrap is incorrect in large samples whenever $n^{1/2}\overline{X}_n$ is positive (which happens with probability $\Phi(-c)$).

- By the same argument, we can't rescue the bootstrap by setting $\theta(\hat{F}_n) = \overline{X}_n$. The parametric bootstrap also fails (with $X_i^* \sim \mathcal{N}(\hat{\theta}_n, 1)$), and in this case, the asymptotic statements hold in finite samples.

The cause of the problem is that $J_\infty(t, F)$ is not continuous in F: consider a sequence $F_n = \mathcal{N}(\theta_n, 1)$ with $\theta_n \downarrow 0$. Then $\max_x |F_n(x) - \Phi(x)|_\infty \to 0$, but $J_\infty(\cdot, F_n)$ is standard normal, while $J_\infty(\cdot, \Phi)$ is the distribution of $\max\{0, Z\}$. ⊠

- The broad lesson is that whenever possible, one should try to choose R_n to be "as pivotal as possible"—not only to obtain asymptotic refinements, but for overall approximation quality.
- The narrow lesson is that the bootstrap will not work when applied naïvely in problems in which boundary issues are important, such as moment inequality problems.

Example 6 (Heavy tails). The nonparametric bootstrap fails for estimating the mean of distributions with fewer than two moments (as does the usual approach based on CLT). If the distribution is parametric, the parametric bootstrap can be made to work. ⊠

Example 7 (Sample maximum). Another classic counterexample due to Bickel and Freedman (1981) is estimation of a sample maximum. Let $X_i \sim F$ with support $[0, \theta]$, let $\hat{\theta}_n = X_{(n)}$ be the sample maximum, and $R_n = n(\hat{\theta}_n - \theta(F))$. The nonparametric bootstrap analog is $R_n^* = n(X_{(n)}^* - \hat{\theta}_n)$. Now, $P(R_n^* = 0 \mid \mathbf{X}_n) = 1 - (1 - 1/n)^n \to 1 - e^{-1}$ as $n \to \infty$. However, $J_\infty(t, F) = 1 - e^{tf(\theta)}$, where $f(x)$ is the density of F (if X_i are $U[0,1]$, this limiting distribution is the exponential distribution). Hence, $P(R_n = 0) \to 0$ and the bootstrap is inconsistent. More generally, the bootstrap tends to fail if the support of $\theta(F)$ determines the support of X_i.

The parametric bootstrap is consistent. ⊠

Example 8. Instrumental variables model with weak instruments, and more generally, weakly identified models. ⊠

Example 9 (Matching estimator). Here bootstrap fails even though the estimator is asymptotically normal: the issue is that the estimator is not asymptotically linear.

Suppose we observe $\{Y_i, D_i, X_i\}_{i=1}^n$, where Y_i is an outcome, D_i a treatment indicator, and X_i a vector of covariates. We're interested in estimating the average treatment effect for the treated under the assumption that the treatment is as good as randomly assigned conditional on covariates. A matching estimator (with a single match) estimates the untreated outcome $Y_j(0)$ for each treated individual with the outcome Y_i of the untreated individual i who is closest to j in terms of covariate distance $d(X_i, X_j)$. The estimator takes the form $N_1^{-1} \sum_i (D_i - (1 - D_i)K_i)Y_i$, where N_1 is the number of treated people, and K_i is the weighted number of times an untreated individual i is used as a match (if two units are closest, both are used as a match with weight $1/2$). Abadie and Imbens (2008) show that the bootstrap fails to reproduce the distribution of K_i. The intuition Abadie and Imbens (2008) give is that if the number of treated units is small, most control units are matched no more than once. But in the bootstrap world, a treated unit can appear multiple times, in which case the control is matched more than once. Otsu and Rai (2017) propose a weighted bootstrap modification that works. Adusumilli (2023) proposes a bootstrap procedure for the propensity score matching estimator that's consistent. ⊠

4. SUBSAMPLING

An alternative way of approximating $J_n(\cdot, F)$ is to use subsamples of size $m < n$ from the original data. These methods often work even in cases in which the bootstrap fails. However, they tend to be less accurate than the bootstrap when it does work. A good summary of these alternative methods is in Horowitz (2001, Section 2.2). Politis, Romano, and Wolf (1999) is the standard technical reference. Much of the theoretical underpinning is due to Wu (1990) and Politis and Romano (1994) and Bickel, Götze, and van Zwet (1997).

The first alternative is called *replacement subsampling*, *m out of n bootstrap*, or *rescaled bootstrap*. It works just like the nonparametric bootstrap in that we draw i.i.d. from $\mathbf{X}_n$, but we only draw a sample of size $m < n$, so that we approximate $J_n(\cdot, F)$ by $J_m(\cdot, \hat{F}_n)$ rather than $J_n(\cdot, \hat{F}_n)$.

Intuitively, the reason this works is that if m is small relative to n, then the sampling error in $\hat{F}_n$ is negligible relative to sampling variability in the bootstrap subsample, making the m out of n bootstrap less sensitive to continuity of J_n in F.

Example 5 (continued). Consider the parameter at a boundary example. The m out of n bootstrap uses the distribution of $m^{1/2}(\hat{\theta}_m^* - \hat{\theta}_n)$ to approximate the distribution of $n^{1/2}(\hat{\theta}_n - \theta)$. We have

$$\begin{aligned}\sqrt{m}(\hat{\theta}_m(\mathbf{X}_m^*) - \hat{\theta}_n(\mathbf{X}_n)) &= \max\{m^{1/2}\overline{X}_m^* - m^{1/2}\theta, -m^{1/2}\theta\} - m^{1/2}(\hat{\theta}_n - \theta) \\ &= \max\{\sqrt{m}(\overline{X}_m^* - \overline{X}_n) + \sqrt{m}(\overline{X}_n - \theta), -m^{1/2}\theta\} - \sqrt{m}(\hat{\theta}_n - \theta).\end{aligned}$$

By the law of iterated logarithm, $\limsup_n \sqrt{n}(\overline{X}_n - \theta)/\sqrt{\log\log n} \overset{\text{a.s.}}{\to} \sqrt{2}$, so that if $\log\log n \cdot m/n \to 0$, then with probability one,

$$m^{1/2}(\hat{\theta}_m(\mathbf{X}_m^*) - \hat{\theta}_n(\mathbf{X}_n)) = \max\{m^{1/2}(\overline{X}_m^* - \overline{X}_n) + o(1), -m^{1/2}\theta\} - o(1).$$

Now, by the central limit theorem if $m \to \infty$, $m^{1/2}(\overline{X}_m^* - \overline{X}_n) \Rightarrow Z \sim \mathcal{N}(0,1)$, conditional on $\mathbf{X}_n$, so that conditionally on $\mathbf{X}_n$, $m^{1/2}(\hat{\theta}_m(\mathbf{X}_m^*) - \hat{\theta}_n(\mathbf{X}_n))$ converges to Z if $\theta > 0$ and to $\max\{Z, 0\}$ if $\theta = 0$ as required. In summary, if $m \to \infty$ and $m(\log\log n)/n \to 0$, then the m out of n bootstrap is consistent. ⊠

The second alternative is called *(non-replacement) subsampling*. Assume that the root has the form $R_n = \tau_n(\hat{\theta}_n(\mathbf{X}_n) - \theta(F))$, where τ_n is a normalizing constant (usually equal to $\sqrt{n}$). In subsampling, we draw a subsample of size m from $\mathbf{X}_n$ *without replacement*. Formally, enumerate the $N = \binom{n}{m}$ subsets of $\{1, \ldots, n\}$, and let $\hat{\theta}_{m,k}$ correspond the statistic $\hat{\theta}(\mathbf{X}_{m,k})$ based on the kth subset. We approximate $J_n(t, F)$ by

$$\hat{J}_{n,m}(t) = N^{-1}\sum_{k=1}^{N} \mathbb{1}\{\tau_m(\hat{\theta}_{m,k} - \hat{\theta}_n) \leq t\}.$$

The intuition for this method is that $\mathbf{X}_{m,k}$ is a random sample of size m from F. Therefore, the *exact distribution* of $\tau_m(\hat{\theta}_{m,k} - \theta)$ is $J_m(\cdot, F)$, where

$$J_m(t, F) = E[\mathbb{1}\{\tau_m(\hat{\theta}_{m,k} - \theta) \leq t\}].$$

The subsampling distribution replaces θ by $\hat{\theta}_n$, and replaces the population expectation by subsample average. If m/n is small, then the error induced by the first substitution should be negligible relative to the randomness in $\hat{\theta}_{m,k}$ (same intuition as with the m out of n bootstrap). Similarly, if N is large, we should be able to appeal to some sort of law of large numbers so that the subsample average is close to the population expectation.

Theorem 5 (Politis and Romano, 1994, Theorem 2.1). Assume that the distribution $J_n(\cdot, F)$ converges weakly to some limiting distribution $J_\infty(\cdot, F)$, and that $\tau_m/\tau_n \to 0$ and $m \to \infty$ as $n \to \infty$. Then:

1. *Subsampling is consistent in the sense that $\hat{J}_{n,m}(t) \xrightarrow{p} J_\infty(t, F)$ at all continuity points t of J_∞ (and hence by an application of Pólya's Theorem, if J_∞ is continuous, then $\|\hat{J}_{n,m} - J_\infty\| \xrightarrow{p} 0$)*

2. *If J_∞ is continuous at its $1-\alpha$ quantile, then the one-sided subsampling CI is asymptotically valid,*
$$P_F(\tau_n(\hat{\theta}_n - \theta(F)) \leq \hat{J}_{m,n}^{-1}(1-\alpha)) \to 1-\alpha.$$

So with subsampling, we don't need to worry about continuity of J_n in F (at least as far as pointwise asymptotics are concerned).

- Because the distribution $J_\infty(\cdot, F)$ is well-defined for all F in Examples 5 to 7 and 9, subsampling will still work (as long as we know the rate of convergence). For instance, in Example 7, $P(R_m^* = 0 \mid \mathbf{X}_n) = 1 - (1-1/n)^m \approx 1 - e^{-m/n} \to 0$ if $m/n \to 0$, and the point-mass problem goes away. Consistency of subsampling in Example 6 was shown in Romano and Wolf (1999).

- By the same logic, subsampling still works in the maximum score example (Delgado, Rodríguez-Poo, and Wolf 2001). Although the limit distribution is not standard, all we need to know that it exists, and that the rate of convergence is $n^{-1/3}$ (and that the limit distribution has no mass points, which they fail to show).

- Subsampling (as well as the m out of n bootstrap) *fails* in Example 8—see Andrews and Guggenberger (2010). Here $\tau_n = 1$, so the condition $\tau_m/\tau_n \to 0$ fails.

On the other hand, choosing m appropriately can be hard in practice. Furthermore, it can be shown that in many cases in which the bootstrap works, it is also more accurate: the bootstrap error is at most $O_p(n^{-1/2})$ versus $O_p(m/n + m^{-1/2})$ for subsampling, which is at least as big as $O_p(n^{-1/3})$ (see Section 2.4 in Politis and Romano 1994), with the optimal choice of m being $m = O_p(n^{2/3})$. This makes subsampling attractive mostly just in situations in which the bootstrap fails (and we don't have available a bootstrap modification that works).

REFERENCES

Abadie, Alberto, and Guido W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76, no. 6 (November): 1537–1557. https://doi.org/10.3982/ECTA6474.

Abrevaya, Jason, and Jian Huang. 2005. "On the Bootstrap of the Maximum Score Estimator." *Econometrica* 73, no. 4 (July): 1175–1204. https://doi.org/j.1468-0262.2005.00613.x.

Adusumilli, Karun. 2023. "Bootstrap Inference for Propensity Score Matching." Working paper, University of Pennsylvania, November. https://www.dropbox.com/s/e4n2wct32uopsyi/PSM-Bootstrap-10.pdf.

Andrews, Donald W. K. 2000. "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space." *Econometrica* 68, no. 2 (March): 399–405. https://doi.org/10.1111/1468-0262.00114.

Andrews, Donald W. K., and Patrik Guggenberger. 2010. "Applications of Subsampling, Hybrid, and Size-Correction Methods." *Journal of Econometrics* 158, no. 2 (October): 285–305. https://doi.org/10.1016/j.jeconom.2010.01.002.

Andrews, Isaiah, and Jesse Shapiro. 2024. *Bootstrap Diagnostics for Irregular Estimators.* Technical report 32038. Cambridge, MA: National Bureau of Economic Research, January. https://doi.org/10.3386/w32038.

Beran, Rudolf. 1984. "Bootstrap Methods in Statistics." *Jahresbericht der Deutschen Mathematiker-Vereinigung* 86, no. 1 (January): 14–30.

Bickel, Peter J., and David A. Freedman. 1981. "Some Asymptotic Theory for the Bootstrap." *The Annals of Statistics* 9, no. 6 (November): 1196–1217. https://doi.org/10.1214/aos/1176345637.

Bickel, Peter J., Friedrich Götze, and Willem R. van Zwet. 1997. "Resampling Fewer than n Observations: Gains, Losses, and Remedies for Losses." *Statistica Sinica* 7, no. 1 (January): 1–31. https://www.jstor.org/stable/26432490.

Cattaneo, Matias D., Michael Jansson, and Kenichi Nagasawa. 2020. "Bootstrap-Based Inference for Cube Root Asymptotics." *Econometrica* 88 (5): 2203–2219. https://doi.org/10.3982/ECTA17950.

de Angelis, D., Peter Hall, and G. A. Young. 1993. "Analytical and Bootstrap Approximations to Estimator Distributions in L^1 Regression." *Journal of the American Statistical Association* 88, no. 424 (December): 1310–1316. https://doi.org/10.1080/01621459.1993.10476412.

Delgado, Miguel A., Juan M. Rodríguez-Poo, and Michael Wolf. 2001. "Subsampling Inference in Cube Root Asymptotics with an Application to Manski's Maximum Score Estimator." *Economics Letters* 73, no. 2 (November): 241–250. https://doi.org/10.1016/S0165-1765(01)00494-3.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7, no. 1 (January): 1–26. https://doi.org/10.1214/aos/1176344552.

Hahn, Jinyong. 1995. "Bootstrapping Quantile Regression Estimators." *Econometric Theory* 11, no. 1 (February): 105–121. https://doi.org/10.1017/S0266466600009051.

Hahn, Jinyong, and Zhipeng Liao. 2021. "Bootstrap Standard Error Estimates and Inference." *Econometrica* 89, no. 4 (July): 1963–1977. https://doi.org/10.3982/ECTA17912.

Hall, Peter. 1988. "Rate of Convergence in Bootstrap Approximations." *The Annals of Probability* 16, no. 4 (October): 1665–1684. https://doi.org/10.1214/aop/1176991590.

———. 1992. *The Bootstrap and Edgeworth Expansion.* New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4384-7.

Hall, Peter, and Joel L. Horowitz. 1996. "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators." *Econometrica* 64, no. 4 (July): 891–916. https://doi.org/10.2307/2171849.

Higgins, Ayden, and Koen Jochmans. 2024. "Bootstrap Inference for Fixed-Effect Models." *Econometrica* 92, no. 2 (March): 411–427. https://doi.org/10.3982/ECTA20712.

Horowitz, Joel L. 2001. "The Bootstrap." Chap. 52 in *Handbook of Econometrics,* edited by James J. Heckman and Edward E. Leamer, 5:3159–3228. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4412(01)05005-X.

———. 2019. "Bootstrap Methods in Econometrics." *Annual Review of Economics* 11 (August): 193–224. https://doi.org/10.1146/annurev-economics-080218-025651.

Kitagawa, Toru. 2015. "A Test for Instrument Validity." *Econometrica* 83, no. 5 (September): 2043–2063. https://doi.org/10.3982/ECTA11974.

Kosorok, Michael R. 2008. "Bootstrapping the Grenander Estimator." In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen,* edited by N. Balakrishnan, Edsel A. Peña, and Mervyn J. Silvapulle, 282–292. Beachwood, OH: Institute of Mathematical Statistics. https://doi.org/10.1214/193940307000000202.

Lehmann, Erich L., and Joseph P. Romano. 2005. *Testing Statistical Hypotheses.* 3rd ed. New York, NY: Springer. https://doi.org/10.1007/0-387-27605-X.

Mammen, Enno. 1992. *When Does Bootstrap Work?* New York, NY: Springer. https://doi.org/10.1007/978-1-4612-2950-6.

Otsu, Taisuke, and Yoshiyasu Rai. 2017. "Bootstrap Inference of Matching Estimators for Average Treatment Effects." *Journal of the American Statistical Association* 112, no. 520 (October): 1720–1732. https://doi.org/10.1080/01621459.2016.1231613.

Politis, Dimitris N., and Joseph P. Romano. 1994. "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *The Annals of Statistics* 22, no. 4 (December): 2031–2050. https://doi.org/10.1214/aos/1176325770.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf. 1999. *Subsampling.* New York, NY: Springer-Verlag. https://doi.org/10.1007/978-1-4612-1554-7.

Romano, Joseph P., and Michael Wolf. 1999. "Subsampling Inference for the Mean in the Heavy-Tailed Case." *Metrika* 50, no. 1 (November): 55–69. https://doi.org/10.1007/s001840050035.

Sen, Bodhisattva, Moulinath Banerjee, and Michael Woodroofe. 2010. "Inconsistency of Bootstrap: The Grenander Estimator." *The Annals of Statistics* 38, no. 4 (August): 1953–1977. https://doi.org/10.1214/09-AOS777.

van der Vaart, Aad. 1998. *Asymptotic Statistics.* New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511802256.

Wu, Chien F. J. 1990. "On the Asymptotic Properties of the Jackknife Histogram." *The Annals of Statistics* 18, no. 3 (September): 1438–1452. https://doi.org/10.1214/aos/1176347759.