607
Methods for applying the bootstrap principle

Jon Faust

http://e105.org/e607

October 13, 2016

- **Bootstrap Principle**
  - Want to know sampling properties of $\phi(Y)$ where $Y \sim P_\theta$.
  - Pick θ_b
  - Take the exact distribution under θ_b as proxy for distribution under the unknown θ^*.
- **A slight sharpening**
  - The above statement is nice and compact and conveys the main idea.
  - But it glosses over an important issue
- **Less compact/more complete:**
  - Model: $Y \sim P_\theta, \theta \in \Theta$
  - Pick a base statistic, e.g.,

    For example, $\phi(Y)$; $\phi(Y) - \phi^*$, or $\sqrt{T}(\phi(Y) - \phi^*)/a\hat{s}e(\hat{\phi}(Y))$, where ase is asymptotic standard error.
  - Pick a θ^b with implied ϕ^b.
  - Use the exact distribution of the base statistic under θ_b as your proxy for the unknown distribution of the base statistic
- **This lecture is about natural applications of this principle**
- **This lecture is not about...**
  - Whether these natural applications work well in any given context
  - Two issues will often be central to the usefulness in any given context.

- Can we pick θ_b well?

  See the lecture on bootstrap DGPs.

- Is the underlying statistic pivotal (or can we implicitly create a pivotal version)?

  See the lecture on The bootstrap and asymptotic refinements.

- This lecture is just a recipe book

  Go elsewhere to see if the implied food tastes good.

▶ **Applications**

- Estimating bias
- Estimating variance-covariance matrices.
- Computing p-values
- Creating confidence intervals

  Those last two are obviously intimately related.

▶ **Notation throughout**

- $Y \sim P_\theta, \theta \in \Theta$
- Truth: θ^*
- Generic scalar statistic: $\phi(Y)$, more compactly we also write $\hat{\phi}$.
- True value of parameter: ϕ^*

  ϕ^* is a function of θ^*

▶ **Bias**

- Want to estimate the bias in $\hat{\phi}$.
- Base statistic: $\Delta = \hat{\phi} - \phi^*$.
- Want to know $E_{\theta^*}\Delta$.
- We'll proxy this by $E_{\theta_b}\Delta$.

▶ **Bootstrap for bias**

- In practice: Draw a zillion samples; compute $\hat{\phi}$ on each; take the mean of these zillion, call it $\bar{\phi}_b$.
- Bias estimate is $\bar{\Delta}_b = \bar{\phi}_b - \phi_b$

▶ **Bias adjusted estimate**

- A natural bias-adjusted estimate is then,

$$\tilde{\phi} = \hat{\phi} - \bar{\Delta}_b$$

Where $\hat{\phi}$ is the original estimate on the sample at hand.

- Note: using the definiton of Δ you can see that,

$$\tilde{\phi} = 2\hat{\phi} - \bar{\phi}_b$$

this formula seems weird to many folks the first time they see it, but perhaps less so the second or third time...

► **Variance-covariance matrix estimation.**

- We did an example of this in discussing White standard errors in the Bootstrap introduction lecture.
- This idea is trivial.
- Base statistic is an estimator $\hat{\phi}$ that is in the CAN framework
- Want to know variance of this statistic: $E_{\theta^*}\hat{\phi}\hat{\phi}'$
- Bootstrap principle: Proxy the unknown item of interest by

$$E_{\theta_b}\hat{\phi}\hat{\phi}'$$

- In practice: Draw a zillion samples according to θ_b; compute $\hat{\phi}$ on each; take the sample variance-covariance matrix as your estimate.

► **p-values/critical values**

- Test is say, $g(\theta) = 0$.
- Base statistic is some test statistic, $\phi(Y, \Theta_0)$
- This notation is meant to emphasize that the test statistic often involves some value defined in the null hypothesis.

For example, the t statistic for $H_0 : \beta = \beta_0$ has $\hat{\beta} - \beta_0$ in the numerator.

- And assume that the statistic is asymptotically pivotal.
- We want to know the quantiles of $\hat{\phi}$ when the null is true. That is, when $\theta^* \in \Theta_0$

Since the stat. is asymptotically pivotal, asymptotically at least (in the first-order asymptotic sense), it doesn't matter which $\theta \in \Theta_0$ is true.

- Bootstrap idea: pick θ^b, and proxy the unknown items of interest by the quantiles of $\phi(Y, \theta_b)$.

- In practice: Draw a zillion samples according to θ_b; compute $\phi(Y, \Theta_0)$ on each; use the quantiles of the EDF of the simulated ϕs as your proxy

▶ **Aside:: Practical detail on critical values and p values.**

- Call the EDF of your simulated statistics EDF_ϕ.
- For statistics that reject for large values, you can pick a nominal size α critical value as the c for which $1 - EDF_\phi(c) = \alpha$
- Suppose that the value of the statistic on the sample at hand is $\hat{\phi}$. Evaluate a the nominal p-value as $1 - EDF_\phi(\hat{\phi})$.
- (One-tailed tests or tests that reject for small values require obvious modifications. . .)

▶ **Confidence intervals: a bit more subtle**

▶ **Confidence intervals**

- We want to know a confidence interval for a scalar parameter ϕ with true value ϕ^*.
- Subtelty 1: Any base test statistic can imply a confidence interval

  Thus, we have a few choices

- Subtelty 2: if the distribution of our base statistic is symmetric (seldom), or asymptotically symmetric (often), we can choose whether or not to exploit this symmetry.

  this may seem puzzling, but should become more clear below.

▶ **3 Confidence interval procedures**

- We have 3 widely used approaches based on these ideas.

▶ **Three CIs**

- Percentile-t bootstrap confidence interval.

  this is based on a studentized test statistic:

$$\sqrt{T}(\hat{\phi} - \phi^*)/a\hat{s}e(\hat{\phi})$$

  .

- Percentile (or Hall's percentile) confidence interval.

  based on the distribution of $\hat{\phi} - \phi^*$.

- The 'other percentile' (or Efron's percentile)

  Like Hall's but (as will become clear below) reverses the treatment of the tails. Under symmetry this may be sensible (one tail is as good as another).

- **Percentile t bootstrap**
  - Supposing we have a CAN estimator $\hat{\phi}$ and a null $\hat{\phi} = \phi^*(\theta)$. then it is natural to consider the base statistic,

$$\tau(Y, \phi^*) = \frac{\sqrt{T}(\hat{\phi} - \phi^*)}{a\hat{s}e(\hat{\phi})}$$

  - If we knew the exact distribution, we could form a confidence interval in a standard way

    Tests generally imply confidence intervals (see the lecture on estimators, tests, and confidence intervals).

- **Aside:: The natural confidence interval**
  - You should know what 'the natural confidence interval' is.
  - This is a generic issue, unrelated to the bootstrap.
  - But in this context, folks get confused.

    So we'll expand on this a bit

- **Aside::**
  - Here we will talk of exact distributions, later sticking in our bootstrap proxy is trivial.
  - The key here will be that the upper tail of the distribution of τ will determine the lower limit on the confidence interval, and vice versa.
  - Define $q(\alpha)$ to be the 100α percent quantile of $\tau(Y, \theta^*)$

    That is, $\mathrm{pr}(\tau < q_\alpha) = \alpha$.

  - We have:

$$\begin{aligned}
\mathrm{pr}(\tau > q_\alpha) &= \mathrm{pr}(\sqrt{T}(\hat{\phi} - \phi^*)/a\hat{s}e > q(\alpha)) = 1 - \alpha \\
\mathrm{pr}(hat\phi - \phi^* > q(\alpha)(a\hat{s}e/\sqrt{T}) &= 1 - \alpha \\
\mathrm{pr}(-\phi^* > \hat{\phi} - q(\alpha)(a\hat{s}e/\sqrt{T}) &= 1 - \alpha \\
\mathrm{pr}(\phi^* < \hat{\phi} - q(\alpha)(a\hat{s}e/\sqrt{T}) &= 1 - \alpha
\end{aligned}$$

- **The percentile-t confidence interval with confidence level** $100(1 - \alpha)$
  - The lower and upper bounds for the percentile-t confidence interval are:

$$\begin{aligned}
c_\ell &= \hat{\phi} - q(\ell)(a\hat{s}e/\sqrt{T}) \\
c_u &= \hat{\phi} - q(u)(a\hat{s}e/\sqrt{T})
\end{aligned}$$

- Where
  - The $q(\ell)$ and $q(u)$ are quantiles of the distribution of the studentized statistic, τ
  - In these qs, $\ell = 100(1-\alpha/2)$ and $u = 100\alpha/2$
  - $a\hat{s}e$ is the estimate of the asymptotic standard error used to studentize.

▶ **Concretely, a 95 percent equi-tailed CI**

- Thus, for a 95 percent equi-tailed confidence interval:

$$\begin{aligned} c_\ell &= \hat{\phi} - q(97.5)(a\hat{s}e/\sqrt{T}) \\ c_u &= \hat{\phi} - q(2.5)(a\hat{s}e/\sqrt{T}) \end{aligned}$$

▶ **Aside:: Staying oriented**

- If τ has a symmetric Gaussian distribution, the confidence interval just suggested implies that we use $\hat{\phi} \pm 2\sigma$ as a 95 percent confidence interval

  Where we have taken the liberties that $1.96 \approx 2$ and that the finite sample standard error, σ, is $a\hat{s}e/\sqrt{T}$.

- Also remember that the studentized statistic is centered on zero so that left tail quantiles are negative values.

  Otherwise, you might think that the upper value falls below the point estimate.

- Notice that the upper tail of the distribution of the estimator ϕ determines the lower bound of the confidence interval and vice versa.

  Under symmetric distributions of the test statistic, this fact has no important implication.

▶ **Percentile confidence interval**

- Next comes the percentile confidence interval, which we might also call Hall's percentile confidence interval
- Follow the logic above, but start with the base statistic $\gamma = \hat{\phi} - \phi^*$

  That is, we skip the studentizing step—we don't divide γ by its standard error—and use this non-pivotal quantity to form our confidence interval.

▶ **The percentile confidence interval**

- Following the logic from above, the $100(1-\alpha)$ percent confidence interval is:

$$\begin{aligned} c_\ell &= \hat{\phi} - q(\ell) \\ c_u &= \hat{\phi} - q(u) \end{aligned}$$

- Where
  - $q(u)$ and $q(\ell)$ are quantiles of the distribution of $\gamma = \hat{\phi} - \phi^*$
  - In the qs, $\ell = 100(1 - \alpha/2)$ and $u = 100\alpha/2$

▶ **Aside:: Why do this?**

- This is mainly a lecture about recipes, not the merits of the recipes
- But you might wonder why you would use the percentile interval rather than the percentile t.
- Main case is when the standard error cannot be stably estimated

  That is, when $a\hat{s}e$ tends to be an erratic estimate of the true asymptotic standard error.

- (Remember, we assume that the estimate of the asymptotic standard error is consistent, so we mean erratic in relevant sample sizes—consistency implies nothing in finite samples.)
- A classic example of when the asymptotic standard error cannot be stably estimated is when the coefficient we are testing is a correlation.

  So that the t statistic would involve a sample correlation over its estimated standard error.

▶ **The 'other percentile' confidence interval**

- If the distribution of our base statistic is symmetric (which is rare) or asymptotically symmetric (which is fairly common) then the distinction about which tail of the statistic determines which bound on the CI is moot.
- That is, for the lower bound of the 95 percent confidence interval, we could equally well use:

$$
\begin{aligned}
c_\ell &= \hat{\phi} - q(97.5) \\
c_\ell &= \hat{\phi} - (q(97.5) - q(2.5))/2 \\
c_\ell &= \hat{\phi} + q(2.5)
\end{aligned}
$$

  because $q(\alpha) = -q(1 - \alpha)$.

- The 'other percentile confidence interval,' also known as Efron's percentile confidence interval, flips the treatment of the tails relative to the percentile confidence interval.

▶ **Other percentile confidence interval**

- The other percentile confidence interval:

$$
\begin{aligned}
c_\ell &= \hat{\phi} - q(u) \\
c_u &= \hat{\phi} - q(\ell)
\end{aligned}
$$

- Where
  - the $q(\ell)$ and $q(u)$ quantiles of $\hat{\phi} - \phi^*$, and
  - $\ell = 100(1-\alpha)$ and $u = 100\alpha$.

▶ **Aside:: Huh? Why?**

- Hey, this is just a recipe.
- You should understand, however, that this is not crazy when $\hat{\phi} - \phi^*$ is close to symmetric around zero.
- And when it seems more crazy, we might figure out a way to fix up the craziness

  (see, e.g., iterated bootstrap and bias corrected bootstrap.)

▶ **Leaving exact land**

▶ **Wrapping up**

▶ **What did we do?**

- In each case discussed above, we pick a base statistic
- And proxy its distribution by the EDF of that statistic under θ_b.
- The only tricky bit is choosing the base statistic.

  $\hat{\phi}$, $\hat{\phi} - \phi$, $\sqrt{T}(\hat{\phi} - \phi)/a\hat{s}e(\hat{\phi})$.

- Choose wisely. . . .

▶ **A more detailed look at confidence intervals for impulse responses in VARs**

▶ **Why?**

- These lecture notes are designed to provide a high level overview of issues that will help you to dig deeper.
- But VAR inference plays a major role in macro.
- It also brings up alost every problem we have spoken about.
- And it probably deserves a somewhat more thorough treatment.

▶ **VARs**

- A VAR is

$$A(L)y_t = \varepsilon_t$$

where y and ε are $(p \times 1)$.

- We generally assume stationarity so that there exists $B(L) = A(L)^{-1}$ and

$$y_t = B(L)\varepsilon_t$$

- The $(i,j)^{th}$ element of B is a scalar lag polynomial,

$$b_{ij}(L) = \sum_{k=0}^{\infty} L^k b_{ijk}$$

- The coefficients of $b_{ij}(L)$ are the impulse response of the i^{th} variable to the j^{th} shock.
- That is, b_{ijk} is the response at lag k of the i^{th} variable to the j^{th} shock.
- Inference about impluse responses raise many tricky issues.

▶ **The natural parametric bootstrap**

- We could use the VAR generalization of the univariate AR(1) parametric bootstrap just described.
- Estimate the VAR, get the $T \times k$ matrix of $\hat{\varepsilon}$s

  (where k is the number of equations in the VAR)

- Recursivly generate new samples; estimate the VAR on each re-sample; compute the implied impules responses on each.
- That is we get a zillion $\hat{A}(L)$s and compute the corresponding zillion $\hat{B}(L)$s
- Now we have a zillion replicated of each b_{ijk} and can proceed as we usually do to compute the bias or standard deviation of the estimates.
- There are two problems.

▶ **Problem 1**

- All our estimators in time series are biased.
- But when persistence lasts a substantial share of the span of the sample, the estimated persistence is strongly biased downward

  We'll get to the basis of this later in the course.

- Thus, when you estimate $\hat{A}(L)$ in persistent data, you underestimate the persistence.

  e.g. in the AR(1) case, the nearer ρ is to 1, the greater the downard bias.

- Then if you use this $\hat{A}(L)$ as the parameters in your parametric bootstrap, you end up using a bad proxy for truth.

- Kilian showed that this is a big deal in macro-type DGPs, but can be substantially ameliorated by usig a bias adjusted $\hat{A}(L)$ as the bootstrap DGP.

  That is, do the bias adjustment described in the intro lecture to form $\tilde{A}(L) = \hat{A}(L) - B\hat{ias}_b$, where that subscript b just signifies a bootstrap estimate of the bias.

- Thus, it is common to do the standard parametric bootstrap, but to use bias adjusted VAR coefficients in the bootstrap.

▶ **Problem 2.**

- On some of the re-samples, you may estimate a VAR the coefficients of which imply nonstationarity.

  The equivalent of estimate $|\hat{\rho}| > 1$ in the univariate AR(1)

- But in this case, you can't compute the impulse response $\hat{A}(L)^{-1}$

  nonstationary means can't invert.

- Let's suppose that the maintained model is that the process is stationary.
- This means that in a sufficiently large sample, even our resamples will give rise to stationary estimates of the VAR with probability near 1.
- Thus, this is a finite sample problem.
- Once again, our whole justification for using the bootstrap is a large sample justification,

  Thus, any reasonable fix to finite sample problems will be fine in the sense of not affecting the large sample merits of the approach.

- For example, whenever we find a $\hat{A}(L)$ on a re-sample that is nonstationary, we could simply replace that $\hat{A}(L)$ with any arbitrary fixed (default value) for $A(L)$.
- Kilian proposes a more attractive solution which you can read in his work, but the mian point is that about anything will do from an asymptotic perspective.
- If you find yourself needing this fix on a very large share of the re-samples, it probably means that your sample is not large enought to trust the large sample justification for the technique.

▶ **Sketch of Kilian's full recommended approach**

- 1. Estimate VAR to get $\hat{A}(L)$
- 2. Run a bootstrap to estimate and apply a bias adjustment to form $\tilde{A}(L)$
- 3. Now run a parametric bootstrap using $\tilde{A}(L)$ and resampling from the EDF of the residuals
- 4. On each draw, compute $\hat{A}(L)_b$; bias adjust it using same Ψ as above to give $\tilde{A}(L)$; invert it to give $\tilde{B}(L)$.

- 5. Form an other percentile confidence interval using the zillion $\tilde{B}(L)$s.
- But also note some details.
- In any bootstrap sample, if you estimate a nonstationary VAR, adjust the estimated $A(L)$ in about any systematic way you like to bring it into the stationary region (and then skip the bias adjustment)

  This step is arbitrary and unimportant asymptotically.

- You could run a nested bootstrap to bias adjust each bootstrap draw; rather than simply using the initial bias adjustment on each draw.

  This is very costly computationally and may not help much, but worth remembering.

- You could compute a percentile or percentile t confidence interval, instead of an other percentile interval.
- These others don't seem to work as well.

  This is just an empirical regularity across the Monte Carlos the profession has examined to date.

  See the Kilian articles cited below.

▶ **Kilian cites**

- Lutz Kilian, 1999. "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses," The Review of Economics and Statistics, MIT Press, vol. 81(4), pages 652-660, November.

  go

  http://www.mitpressjournals.org/doi/pdf/10.1162/003465399558517

- Lutz Kilian, 1998. "Small-Sample Confidence Intervals For Impulse Response Functions," The Review of Economics and Statistics, MIT Press, vol. 80(2), pages 218-230, May.

  go

  http://www.mitpressjournals.org/doi/pdf/10.1162/003465398557465

▶ **Literature**

- Sims and Zha propose a Bayesian alternative that is sometimes used.
- Sims, Christopher A., Zha, Tao, Error Bands for Impulse Responses,Econometrica 67,5,1999,1468–0262

  go

  http://dx.doi.org/10.1111/1468-0262.00071

- Even if you don't like their method, their discussion of the problems with impulse response inference is particularly lucid.