607
Bootstrap DGPs: choosing P_{θ_b}

Jon Faust

http://e105.org/e607

October 13, 2016

▶ **Bootstrap Principle**

- Model, $Y \sim P_\theta, \theta \in \Theta$.
- Want to know sampling properties of $\phi(Y)$ where $Y \sim P_{\theta^*}$.

  θ^* is the unknown true value.
- Pick θ_b
- Take the exact distribution under θ_b as proxy for distribution under the unknown θ^*.

▶ **Choosing θ_b is key**

- Choosing the bootstrap distribution, P_{θ_b}, wisely is obviously a key to making this approach useful
- This lecture is about choosing θ_b which I refer to as choosing a bootstrap DGP

▶ **Choosing θ_b**

▶ **A fully and explicitly parametric option**

- In principle, we could pick θ_b explicitly
- And then figure out how to generate samples according to P_θ.

▶ **A fully parametric option, Nope**

- Our macro models almost always have some unspecified or 'not carefully specified' part (also known as a nonparametric part)
- Formally, we can think of this as $\theta = \{\gamma, \psi\}$ where γ is of interest and carefully specified and ψ is a (potentially infinite-dimensional) nuisance parameter.

- We may not have a good way to choose ψ_b, or may not want to invest in thinking about a good way, or may not want to think about how to generate data according to ψ_b even if we could pick one.
- Under these circumstances, a fully parametric bootstrap DGP is not attractive.
- In most cases, we use either a fully non-parametric or semi-parametric approach approach where we pick a γ_b and deal with ψ nonparametrically.
- I'll start by discussing fully nonparametric approaches.

▶ **Nonparametric options**

▶ **Lots of available choices**

- As with most techniques motivated by large sample theory, once we have 1 good method, we have arbitrarily many

  Recipe: take the 1 good method and add adjustments that matter only in finite samples.

- This lecture simply gives a taste of many 'basic' methods and points you toward natural approaches to elaboration.
- Each can endlessly be elaborated; sometimes these elaborations are really important.

▶ **Intuition for bootstrap DGPs**

- Define DGP_θ to be the DGP implied by P_θ.

  P_θ is a probability measure, DGP_θ is a recipe for generating data according to P_θ.

- Fact: our P_θs generally imply or assume a DGP with some iid (or nearly iid) element at the heart of things.
- Perhaps the data are iid (simplest case).
- Perhaps the data are an ARMA:

$$A(L)y_t = B(L)\varepsilon_t \quad \varepsilon_t \; iid$$

  The time series process for y_t is built up from iid εs.

▶ **Strictly stationary data**

- If the data are strictly stationary, any two blocks of B contiguous observations are identically distributed (by definition)

  but they won't generally be independent.

- But any two blocks of B observations that are from points far enough apart in a large sample will arbitrarily close to independent.
- Thus, we have some identically distributed and nearly independent items (the blocks) that we can use for a nonparametric scheme.

▶ **More elaborate**

- Perhaps the data,X, are independent but heteroskedastic.
- We can think of this x_t as the product of two iid random variables:

$$x_t = z_t^{1/2} u_t$$

where z and u are independent, $u \sim iidN(0,1)$ and z is iid with positive support. The standard deviation of x_t is then z_t.

▶ **Bottom line**

- The DGPs tend to be built up from some iid element.

▶ **Why is this important?**

- Often our re-sampling schemes involve
  - construting sample analogs to the iid element.
  - Then resampling from the EDF of these elements

    (that is, drawing uniformly and with replacement from the collection of sample elements)
- And then we build up the new sample implied by this new draw of the of iid elements.

▶ **Terminology**

- In practice, many of these bootstrap DGPs we consider have the form of 're-sampling schemes'
- That is, we get some collection of items from the sample at hand
- And then our DGP involves (at least in part) repeatedly drawing, or re-sampling, from these items.

▶ **First basic method: the EDF and iid data**

- $Y \sim P_\theta, \theta \in \Theta$.
- Suppose Y is $(T \times K)$ with iid rows (observations)
- Natural nonparametric bootstrap DGP involves generating new samples by drawing observations (rows of Y) with replacement.

- This is equivalent to drawing from the empirical distribution function of the rows of Y.
- We called this the XY scheme when the data matrix has the LHS and RHS variables for a regression.
- With iid data, this approach makes a lot of sense
- Generally, if the data elements are iid $\sim F_{\theta^*}$, then

$$EDF(c) \to F_{\theta^*}(c)$$

  for all c as the sample size gets large.
- There is often little reason to go beyond the EDF in the case of iid data.

  But there are some potentially useful elaborations

▶ Some natural elaborations

- Take the Y $(T \times 1)$, univariate case for simplicity.
- Suppose that your maintained model satisfies some additional restrictions
- Say, that y_t is also mean zero and symmetric
- The EDF is won't exactly satisfy these restrictions.
- Thus, you may want to modify the EDF in some natrual way to impose these constraints.
- Note: in very large samples, the EDF will be very close to mean zero and symmetric, so whatever adjustments you make will be unimportant from the standpoint of conventional asymptotics

▶ Mean zero, symmetric

- Easiest way to impose mean zero and symmetric on an EDF is this:
- 1. Before resampling, substract the mean from the Ys; call the re-centered version $\tilde{Y}$.

  The EDF of $\tilde{Y}$ will be mean zero.
- 2. Then resample uniformly from the elements of $\tilde{Y}$ AND $-\tilde{Y}$
- By re-centering and then drawing from Y and $-Y$ we achieve a mean zero and symmetric distribution.

▶ Other elaborations

- There are other elaborations that are of a similar character.
- Modify the EDF in some way to bring it closer to the maintained properties of the model.

  That is, properties that are not in question.

- In large samples, the EDF will come arbitrarily close to conforming to these properties, but in relevant sample sizes you may want to impose them.

▶ **Time series, nonparametric**

▶ **Time series**

- Nonparametrics are more subtle in time series because there may be no natrual parameter-free elements that are completely iid.
- But suppose that the data are strictly stationary and covariance stationary.

  stronger than needed, but may help intuition.
- Consider the following recipe for re-sampling
- Break the sample into b blocks of B contiguous observations.

  Obviously, we either need $T = bB$ or we have one block left over at the end that has less than B observations. Let's just set aside this niggling issue.
- Build new samples by taking these blocks and drawing them with replacement and sticking them end to end.

▶ **Heuristic argument why this is sensible**

- If the blocks were iid, this would just be a slight generalization of the univariate iid EDF case.
- Since we assumed strictly stationary, the blocks are identically distributed

  That is essentially the definition of strict stationarity (any two such blocks have the same distribution)
- But three related problems:
  - 1. The resampling scheme obviously produces samples with zero autocorrelatoins at any lag longer than the block size.
  - 2. The block boundaries are a problem: the correlation of x_t and x_{t-1} is zero in the re-samples in this scheme whenever the two span a boundary.
  - 3. The blocks are not independent.

▶ **Answers to the 3**

- The answers to the 3 are all large sample answers.
- That is, they are arguments that if we let the sample get large and the block size get large and the number of blocks get small relative to the sample size, then the 3 problems get vanishingly small.

▶ **Problem 1**

- Problem 1 will become trivially small as we make the block size arbitrarily large.

  Dependence must die out due to stationarity. In fact, to do our proofs, we can go a bit further than stationarity and impose a mixing condition that controls how rapidly the dependence dies out.

- Note, this is really no different from approximating an AR by a finite-order MA(p).

  With p finite, the autocorrelations of the MA are zero after lag p and so won't match the AR wich has nonzero autocorrelations at all lags. But, make p big, and what you miss can't be very important.

▶ **Problem 2**

- Problem 2 will become less important as the ratio of total observations to block boundaries gets small.
- If the block size is B then If the number of blocks in the sample is $b = T/B$
- If b grows more slowly than T then block boundaries are becoming increasingly sparse in the sample.

  In the limit, they become so rare as to be unimportant (in the relevant sense)

▶ **Problem 3: the blocks are not truly independent**

- If there are a large number of large blocks, the dependence between any two blocks chosen at random will be small.

  Number the blocks from 1 to b and ask about the likely distance between two numbers chosen uniformly on $1, \ldots, b$,

▶ **End of hand waving**

- Hopefully, you are starting to see how one might prove that our standard EDF logic almost applies to the blocks in this case.
- That is, if B and b are large and b is small relative to T the blocks are almost iid, so resampling from the blocks makes sense.
- Figuring out an elegant proof of this is tricky, but even reading and understanding that proof is not so hard.

▶ **Elaborations**

- There are a number of ways we can refine the block bootstrap
- None help much, but they do help a bit in some contexts.
- Holding fixed both the block size and the block boundaries causes some undesirable properties.
- We cannot avoid block boundaries, but they don't have to come in fixed locations in the original sample.

We could fix the block size, B, but draw the starting point of the block uniformly and then take the next B observations.

- Or we could build samples by repeatedly drawing both draw the block size and the block starting point.

  so long as the typical block size is growing with sample size, this may be a good idea.

- A version of this is called the stationary bootstrap and can result in a DGP with somewhat better properties than the original fixed block plan.

▶ **Nice summary by the creators of stationary bootstrap**

- Politis and Romano, Stationary Bootstrap

  go

  http://www.jstor.org/stable/2290993

▶ **A macro problem**

- We need the block size to be larger than the important dependence in the sample.
- And we need a large number of blocks.
- As I've said repeatedly, our samples in time series are often short relative to the span of dependence we are studying.

  e.g., business cycles last 10 years and we have 30 years of data.

- This suggests that these block techniques will not be useful for our problems

  And, in fact, in many macro contexts, these techniques are rarely used.

- But, so long as *most* of the dependence dies out fairly quickly, we may be ok
- In practice, the block bootstrap seems to perform well in some Monte Carlos calibrated to match macro applications.

  one demonstration of this is in the thesis work of JHU PhD Pierangelo De Pace (now tenured at Pamona)

▶ **Let's move to some semi-parametric time series schemes**

▶ **Semi-parametric DGPs**

- Your maintained model is P_θ, $\theta \in \Theta$
- As noted above, often $\theta = \{\gamma, \psi\}$ where ψ is a potentially infinite-dimensional nuisance parameter.

- Often conditional on the parametric part, the nonparametric part is independent or iid and something we can re-sample using on of our nonparametric schemes.
- In particular, our maintained model may have iid εs driving the processes.
- And we can estimate the model, compute $\hat{\varepsilon}$ and do the usual thing: treat $\hat{\varepsilon}$ as a proxy for ε.
- So we use the EDF of the $\hat{\varepsilon}$ as what we resample from.

▶ **Regression example**

- Take a regression case with fixed Xs,

$$Y = X\beta + \varepsilon$$

  where ε_t is iid.
- Implicitly, this is a semi-parametric problem with the unstated parameter describing the specifics of the distribution of ε.
- We can estimate $\hat{\beta}$ and compute the implied $\hat{\varepsilon}$s.
- We can make new samples of Y as

$$Y = X\hat{\beta} + \tilde{\varepsilon}$$

  where the $\tilde{\varepsilon}$s come from drawing with replacement from the $\hat{\varepsilon}$s.
- Here we are keeping the Xs fixed at there sample values.

▶ **The key**

- Conditional on a parameter choice for the parametric part, we construct a sample version of an iid element (the $\hat{\varepsilon}$)
- And we resample from the EDF of this element.

▶ **Elaborations**

- The $\hat{\varepsilon}$s generally won't match all the features of the εs assumed in the maintained model.
- If ε is mean zero, but the regression has no constant, the EDF of the $\hat{\varepsilon}$s will not be mean zero.
- We may want to 'adjust' the EDF in some way in order to impose aspects of the maintained model.
- For example, resample from re-centered $\hat{\varepsilon}$s

  that is, subtract the sample mean of the $\hat{\varepsilon}$ before re-sampling.

▶ **Time series regression case**

- When the X are not fixed, we have to think about dependence between the ε and the Xs.

Our bootstrap DGP needs to match the true process in this regard.

- Suppose that we have an AR(p) or VAR(p) model and the model treats the εs as iid.
- In this case, we can estimate the model and compute the $\hat{\varepsilon}$ and then recursively generate a new sample.

Of course, as always i AR work, we have to say something about those pesky initial conditions.

- Suppose the DGP is an AR(1) and we are willing to condition (or take as fixed) the first observation.
- We estimate by OLS using observations 2 through T and get $T-1$ $\hat{\varepsilon}$s.
- Then our new samples are recursively generated by

$$y_{t+1} = \hat{\alpha} + \hat{\rho} y_t + e_{t+1}$$

where e_{t+1} is a draw from the EDF of the $\hat{\varepsilon}$s.

- This is a standard bootstrap used in time series.
- Suppose that ρ is large enough that the dependence lasts a signficant share of the total time span of the sample.

Thus, we might worry about the block bootstrap

- So long as we can get a good estimate of ρ this might not be a big problem: our re-samples will match the dependence in the sample.

Note: that 'so long as' is a big deal and we'll comment further on it below.

- This technique can be used in any AR, MA or even VAR or VMA.

Something like this is widely used in time series applications.

▶ The weird and the wild

▶ The Wild bootstrap.

- Just to give one more mind-bending example, let's consider the wild bootstrap often used in cases of heteroskedasticity.
- Suppose we have a set of $\hat{\varepsilon}$s and we want create new samples that in some way match the EDF of these $\hat{\varepsilon}$s.
- Suppose our main concern is matching the first two or three moments of the $\hat{\varepsilon}$s.

Just take this as given. You can read the cited literature to see why this might be sensible.

- One approach is to leave the $\hat{\varepsilon}$s in their original order, but create a new sample of $\hat{\varepsilon}$s by hitting the original $\hat{\varepsilon}$s with some random variable:

$$e_t = \hat{\varepsilon}_t z_t$$

  where e_t is the t^{th} observation in the new sample, and z_t is chosen iid from a distribution that leaves the e_ts having the same first three moments as is implied by the EDF of the $\hat{\varepsilon}$s.
- For example, if the $\hat{\varepsilon}$s are mean 0, then this multiplication won't change the first moment.
- The variance of the es (when $\hat{\varepsilon}$ is mean 0) is:

$$\text{var}(e_t) = Ee^2 - (Ee)^2 = Ez_t^2 E\hat{e}^2 = Ez_t^2 \,\text{var}(\hat{\varepsilon})$$

  So we just need that $Ez^2 = 1$
- Notice that so far, we can still freely choose the mean of z: we do that to match the third moment of the $\hat{\varepsilon}$s.
- See the Hansen text (or do your own derivation) for simple versions of zs that fit the bill

▶ Wild bootstrap

- You may wonder why anybody would use this technique, but you should see that it is techically possible to choose an iid distribution for the z that performs in the desired manner.
- As to the merits of this approach, see The wild bootstrap tamed at last, by Davidson and Flachaire

  go

  http://ideas.repec.org/p/ecm/wc2000/1413.html

▶ Abstract

- –

  In this paper we are interested in inference based on heteroskedasticity consistent covariance matrix estimators, for which the appropriate bootstrap is a version of the wild bootstrap. Simulation results ... show that all wild bootstrap tests exhibit substantial size distortion if the error terms are skewed and strongly heteroskedastic.
- –

  The distortion is however less, sometimes much less, if one uses a version of the wild bootstrap, belonging to a class we call "tamed", which benefit from an asymptotic refinement ... This version always gives better results than the version usually recommended in the literature, and gives exact results for some specific cases.

- –

  However, when exact results are not available, we find that the rate of convergence to zero of the size distortion of wild bootstrap tests is not very rapid: in some cases, significant size distortion still remains for samples of size 100.

▶ **That is long winded, but the bottom line is,**

- The tamed wild bootstrap is better, sometimes much better, but this can still be veryt bad.

▶ **Wrapping up: bootstrap DGPs**

- These schemes all can essentially be thought of as ways to simulate a process that has lots of the features of the real sample you are working with
- Some of the approaches are clever (e.g., the wild bootstrap) but once you see the ideas, they should make sense.

▶ **Asymptotic results**

- The proof of the virtue of the bootstrap is a large sample result
- It essentially requires that θ_b is chosen in such a way that P_{θ_b} gets arbitrarily close in the relevant sense to P_{θ^*}.

  That is, the chosen bootstrap distribution in the limit looks a lot like truth.