Resampling & the Bootstrap

Ethan Ligon

April 22, 2020

The Real World Data-Generating Process$^{\mathrm{TM}}$

Suppose that at any particular moment in time t, we can describe the *state* of our world by a variable $s_t \in S$, and the history of previous states up to t by by $s^t \in S^t$.

Observed Data

Given a particular history s^t, different economic agents observe (possibly different) sets of reported measurements (which censor, select, and may add error):

$$d_t = \mathcal{R}(s^t)$$

The Real World Data-Generating Process$^{\text{TM}}$

Decisions

Given a particular history s^t, economic agents take actions $y_t = \mathcal{M}(\mathcal{R}(s^t))$. The realization y_t becomes part of the next period's state.

History's Evolution

The state of the world in the subsequent period depends on a law of motion:

$$s_{t+1} = \mathcal{F}(s^t, y_t)$$

The RWDGP$^{\mathrm{TM}}$

The RWDGP can thus be described by a triple $(\mathcal{M}, \mathcal{R}, \mathcal{F})$. Initialize with the history to time zero, s^0, and it returns a corresponding dataset.

Interpretation

- So far the RWDGP has produced all the data available to us. This dataset d^t is finite, but depends on the particular history s^t realized up to this point.
- A different history $\tilde{s}^t$ would have produced a different finite dataset $\mathcal{R}(\tilde{s}^t)$.

The RWDGP$^{\text{TM}}$

Resampling the RWDGP:

- We can imagine having the god-like power of "resampling" from the RWDGP, with the nth draw involving a new initial history s_n^0 and delivering a new dataset $d_t^n = \mathcal{R}(s_n^t)$.
- As our draws $n \to \infty$ we can imagine using our collection of different realizations of datasets to draw inferences. In what proportion of draws is there a global pandemic in the early 21st century? In what proportion of draws do fewer than 10% of Californians earn the minimum wage?

Our Monte Carlo Data-Generating Process

Last time we discussed the creation of a DGP that can be described as a triple $(\mathcal{M}, \mathcal{R}, \mathcal{F})$. Feed in a initial state s^0 (e.g., a seed to a pseudo-random number generator) and it returns a dataset $d = (y, X, Z)$.

Interpretation

- We have the god-like power of resampling from our DGP; Each draw from our DGP produces a different finite dataset. Call these Monte Carlo draws.
- There's no limit on the number of draws we can make from the dataset. As our draws $m \to \infty$ we may be able to draw increasingly accurate inferences about $(\mathcal{M}, \mathcal{R}, \mathcal{F})$ (this is a question of identification).

Our Monte Carlo Experiment

Our particular experiment involved repeated draws to explore the finite sample properties of linear IV in the GMM two-step estimator. We found circumstances under with the limiting distribution of b_N was very different from the estimated empirical distribution.

Three different possible takeaways

1. We need a different estimator with better finite sample properties. (Explore this in discussion question).
2. We need more data. Or;
3. We could use the estimated empirical distributed for inference & hypothesis testing.

After you collect your data

Use the empirical MC distribution, and assume that the MC DGP is close enough to the actual real-world DGP that the empirical distribution of β can serve for testing & inference.

Issue

May requires a lot of confidence in the MC DGP. And if you have this much confidence you may want to use the MC DGP to actually help *estimate* the parameters.

Estimating parameters (Indirect Inference)

Idea: Choose "truth" parameters to make simulated data from the Monte Carlo DGP (in this setting called the 'auxiliary model') match moments or distributions observed in the real-world data. Often used when economic model involves parameters which are complicated functions of the data.

Examples

- Method of Simulated Moments (MSM/SMM)
- Maximum Simulated Likelihood (MSL/SML)
- Monte Carlo Integration
- Full Information Maximum Likelihood (limiting case)

Before you collect data

Standard power calculations usually assume a normal model with very limited forms of dependence. But what if your estimated coefficients aren't normally distributed?

- Typically wind up collecting too little data and being under-powered.
- Use MC distribution instead, where the experiment is actually measuring the finite sample properties of the estimator you'll use when you write your dissertation.
- How big a sample do you really need to achieve a given level of power in your MC experiment?

Bootstrap

Issue with Monte Carlo is that we have to construct a model to build estimates. This will often require us to assume more than we wish to about the Real World DGP.

Alternative

Use the RWDGP! We begin by observing a sample of N observations X_j once; say D_N. If these are independent (they're identically distributed by construction) we just need to figure out how to repeat this draw.

Sampling

Since D_N is comprised of N iid assumptions we can use this sample to construct an empirical distribution function of X, say $\hat{F}$. Then think of simply drawing samples from this empirical distribution.

Non-parametric estimator of empirical distribution function

$$\hat{F}(x) = \frac{1}{N} \sum_j \mathbb{1}(X_j \leq x)$$

Simplification

Since the probability of drawing a particular X from $\hat{F}$ is proportional to the frequency with which X appears in D_N, there's an trivial simplification: instead of constructing $\hat{F}$ just:

1. Draw X_j from D_N.
2. Repeat until you have the sample size you want; often (usually?) this will be N, the size of the original sample. Call the resulting "bootstrap" sample D_N^1.

Interesting fact

What's the probability that a given observation in D_N appears in D_N^1?

Basic Bootstrap estimation

Suppose we want to estimate a vector of parameters β. We can construct an estimate of this using the original sample, say b_N. But we may not know much about the distribution of this estimator.

Procedure

1. Having drawn a bootstrap sample D_N^1, use it to estimate b_N^1.
2. Draw a new sample D_N^2, and compute b_N^2
3. ... Repeat M times...
4. Calculate the sample covariance matrix of the estimates of β,

$$\hat{V}_N^M = \frac{1}{M} \sum_m (b_N^m - \bar{b}_N)(b_N^m - \bar{b}_N)^\top$$

5. Repeat: compute additional bootstrap samples until

$$\|\hat{V}_N^M - \hat{V}_N^{M-1}\| < \epsilon$$

Use

We've just described the construction of a covariance matrix for the estimator b_N via the bootstrap, so this can be used for testing and inference in the usual way. But note that the "usual way" assumes that the distribution of b_N is normal.

Non-normal distributions

In finite samples our distributions may be decidedly *non*-normal. But we have an estimate of the distribution! Just construct the empirical distribution of the M bootstrapped estimates of β.

- Tests of normality available
- Simple construction of confidence intervals