14.385
Nonlinear Econometrics

Lecture 6.

Theory: Bootstrap and Finite Sample Inference

Reference: Horowitz, Bootstrap.

Review 381 notes for finite-sample inference in the regression model with non-normal errors.

Setup:

$\{X_i, i \leq n\}$ is data.

$F_0 \in \mathcal{F}$ is DGP.

Statistical model of DGP:

$\mathcal{F} = \{F(x,\theta), \theta \in \Theta\}$ (parametric model).

$\mathcal{F} = \{F$ is a cdf $\}$ (nonparametric model).

Statistic of interest :

$$T_n = T_n(X_1, ..., X_n).$$

Example. $T_n = t_j$ a t-statistic based on $\widehat{\beta}_j$ in linear reg.

$G_n(t, F_0) = P_{F_0}(T_n \leq t)$ exact df.

$G_n(t, F) = P_F(T_n \leq t)$ exact df under F.

We want to estimate $G_n(t, F_0)$ in order to

- use α- quantiles of $G_n(t, F_0)$, denoted

$$G_n^{-1}(\alpha, F_0) = \inf\{t : G_n(t, F_0) \geq \alpha\},$$

for confidence regions and hypothesis testing,

- use the p-values

$$1 - G_n(t, F_0)|_{t=T_n},$$

for testing, when a large value of T_n suggests a rejection.

The asymptotic df under F_0 is

$G_\infty(t, F_0) = \lim_n G_n(t, F_0)$.

and asymptotic df under F is

$G_\infty(t, F) = \lim_n G_n(t, F)$.

Asymptotic Estimation Principle:

Estimate $G_n(\cdot, F_0)$ using $G_\infty(\cdot, F_0)$.

This is the usual principle we use.

Bootstrap Estimation Principle:

Estimate $G_n(\cdot, F_0)$ using $G_n(\cdot, \widehat{F})$.

Two choices for $\widehat{F}$:

(i) $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} \Rightarrow$ nonparametric bootstrap

(ii) $\widehat{F}_0(x) = F(x, \widehat{\theta}) \Rightarrow$ parametric bootstrap.

There are also hybrid versions.

Monte-Carlo Algorithm for Tabulation of $G_n(t, \hat{F})$.

1. For $j = 1, ..., B$ generate a bootstrap sample of size n,

$$\{X^*_{ij}, i = 1, ..., n\},$$

by sampling from $\hat{F}$ randomly. If $\hat{F}$ is an empirical df, sample estimation data randomly with replacement.

2. Compute

$$T^*_{nj} = T_n(X^*_{1j}, ..., X^*_{nj}), \quad j \leq B.$$

3. Use the sample $\{T^*_{nj}, j \leq B\}$ to compute the empirical probability of the event $\{T^*_n \leq t\}$. For estimates of quantiles, simply use the empirical quantiles of the sample $\{T^*_{nj}, j \leq B\}$.

Finite-Sample Principle:

1. Estimate $G_n(\cdot, F_0)$ using $G_n(\cdot, F_0)$ if know F_0. This is generally infeasible, but this is the idea.

2. In some lucky cases, exact distribution of the statistic does not depend on F:

$$G_n(t, F_0) = G_n(t, F) \quad \forall F \in \mathcal{F}.$$

If this holds, then the statistic T_n is said to be **pivotal** relative to the statistical model $\mathcal{F}$. Thus, under pivotality, we can use any DGP $F \in \mathcal{F}$ to tabulate $G_n(t, F_0)$.

Example 1. The t-statistic in Normal Gauss Markov Model with normal disturbances, $\mathcal{F}_{GMN}$, follows a t-distribution

$$T_n = t_j \sim t(n - K).$$

that does not depend on DGP F within this model. That is, the distribution does not depend on the unknown parameters such as the regression coefficient β and variance of disturbances σ^2.

Its distribution $G_n(t, F_0)$ was tabulated by Student.

Example 2. The t-statistic in Gauss Markov Model with t-disturbances with known degrees of freedom θ (denoted by $\mathcal{F}_{GMt(\theta)}$) is pivotal, as we argued in 14.381.

That is, the distribution does not depend on the unknown parameters such as the regression coefficient β and variance of disturbances σ^2.

Denote this statistical model by $\mathcal{F}_{GMt(\theta)}$.

In 381, we *tabulated* its distribution $G_n(t, F_0)$ using Monte-Carlo.

Recall details: in 381 we did this for $n = 6$ motivated by Temin's data-set on Roman wheat prices.

3. (Advanced Finite-Sample Inference) If do not know F_0, put bounds on $G_n(t, F_0)$ using

$$B_n = [\inf_{F \in \widehat{\mathcal{F}}} G_n(t, F), \quad \sup_{F \in \widehat{\mathcal{F}}} G_n(t, F)]. \quad (1)$$

Use them to bound the p-values and quantiles of $G_n(t, F_0)$.

Example 3. Recall Example 2, but now take t-statistic in classical linear regression model with t-disturbance with unknown degrees of freedom θ. Recall that in Temin's exercise, we have constructed $\mathcal{F}$ by varying θ over a set of "reasonable values" for degrees of freedom provided by an expert:

$$\mathcal{F} = \{\mathcal{F}_{GMt(\theta)}, \theta \in \{4, 8, 30\}\}.$$

We then set $\widehat{\mathcal{F}} = \mathcal{F}$, tabulate p-values and quantiles of our statistic under each $F \in \widehat{\mathcal{F}}$, and then we take the least favorable estimates.

Choice of $\widehat{\mathcal{F}}$. Ideally, we want to choose the set $\widehat{\mathcal{F}}$ such that

$$G_n(t, F_0) \in B_n.$$

For this purpose, it suffices but is not necessary that

$$Prob[F_0 \in \widehat{\mathcal{F}}] \to 1.$$

Also we want $\widehat{\mathcal{F}}$ to "converge" to F_0 in the sense that

$$d(B_n, G_n(t, F_0)) \to_p 0,$$

i.e. the distance between B_n and $G_n(t, F_0)$ goes to zero.

This allows us to conduct asymptotically efficient inferences, while preserving finite sample validity.

Remark*: The convergence to a singleton can fail when $F \mapsto G_n(t, F)$ is not continuous in F at $F = F_0$.

In parametric cases it suffices to set

$$\widehat{\mathcal{F}} = \{F(\cdot, \theta) : \theta \in CI_{1-\beta_n}(\theta_0)\}, \quad \beta_n \to 0,$$

where $CI_{1-\beta_n}(\theta_0)$ could be constructed by the usual means.

Computation: Computation of the conservative bound B_n may seem like a laborious task. However, its success depends on finding **at least one** θ' that yields inferences that are more conservative than using θ_0.

Computational Idea (Perturbed Bootstrap):

1. Start with $\theta_1 = \widehat{\theta} \in CI_{1-\beta_n}(\theta_0)$. Tabulate $G_n(\alpha, F(\cdot, \theta_1))$. That is, the first step is just the bootstrap.

2. Draw a nearby value $\theta_2 = \widehat{\theta} + \eta \in CI_{1-\beta_n}(\theta_0)$, where η is some random perturbation. Tabulate $G_n(\alpha, F(\cdot, \theta_2))$.

3. Repeat step 2 until some clear stopping criterion is reached.

4. For purposes of inference take the *least favorable p- or critical value* generated in this way.

5. Save your code and the Monte-Carlo seed for replicability.

We are thus guaranteed to do *at least* as well as in parametric bootstrap.

Applicability. This finite-sample method is applicable under various non-standard conditions, including

– small sample sizes,
– partially identified models,
– non-regular models,
– moment inequalities.

The method is increasingly becoming more feasible with better computing.

Excellent Example: Oleg Rytchkov's Dissertation (Sloan Ph.D. 2007).

(Technical.* Proceed with care.) In non-parametric cases we can set

$$\widehat{\mathcal{F}} = \{F : F \in CI_{1-\beta_n}(F_0)\}, \quad \beta_n \to 0,$$

where the confidence regions $CI_{1-\beta_n}(F_0)$ collects all cdfs F such that

$$F(t) \in [F_n(t) - c_n(1-\beta_n), F_n(t) + c_n(1-\beta_n)],$$

where

$c_n(1-\beta_n)$ is $1-\beta_n$ quantile of the limit distribution of Kolmogorov-Smirnov Statistic

$$\sqrt{n} \sup_t |F_n(t) - F_0(t)|.$$

This limit distribution is given by the distribution of random variable

$$\sup_t |B(t)|,$$

where $B(\cdot)$ is a F_0-Brownian bridge that describes the limit distribution of the random map

$$\sqrt{n}(F_n(\cdot) - F_0(\cdot)).$$

Under F_0 continuous, $B(\cdot)$ is pivotal, and its distribution has been tabulated.