607
Estimators, tests, and confidence intervals: review, preview, and a sketch of deeper issues

Jon Faust

http://e105.org/e607

October 13, 2016

▶ **Readings**

- You should probably all read Engle's Handbook chapter on the trinity of tests

  go

  http://dx.doi.org/10.1016/S1573-4412(84)02005-5

- If you really want to get into the depths of testing, you should spend a good deal of time with Lehmann's book 'Testing Statistical Hypotheses'

  A classic

- Several articles are cited below as well.

▶ **This lecture**

- This lecture reviews, or skates across a wide range of standard terms and concepts relating tests, confidence intervals and estimators.
- We start talking about exact results and then turn to asymptotics
- And in the end we gesture at deeper issues.

▶ **Tests, estimators, and CIs**

- Why do we form point estimates of parameters?

  Well, in part, because we want point estimates, but more importantly, because estimators are often a natural route to what we really need: confidence intervals and tests

- For example, if the estimator is in the CAN framework then

$$\hat{\beta} \pm 1.96\hat{se}(\hat{\beta})$$

  is a valid nominal 95 percent confidence interval.
- And confidence intervals and tests are really two ways of thinking about the same topic.

▶ **Test**

- A classical test is defined by a hypothesis and a rejection region, also called a critical region.

  rejection region is a set of observable outcomes

- The hypothesis is considered rejected if we observe an outcome in the rejection region.

▶ **More specifically**

- $Y \sim P_\theta, \theta \in \Theta$, $Y \in \mathcal{Y}$.
- A hypothesis is given by a subset of θ: $H_0 : \theta \in \Theta_0$.
- A critical region is as subset of $\mathcal{Y}$, $\mathbf{Y}^c$
- Of course, rejection regions are typically defined in terms of statistics and critical values.
- For example:

$$\mathbf{Y}^c = \{Y \in \mathcal{Y} | \phi(Y) > c\}$$

  where c is called the critical value.

▶ **Size**

- The size of the test is defined as

$$\sup_{\theta \in \Theta_0} P_\theta(Y \in \mathbf{Y}^c)$$

- This is the maximum probability of type I error taken over all ways that the null could be true.

  type I error is rejecting the null when it is true.

▶ **Power**

- Power is the probability you reject the null when some particular $\theta_K \notin \Theta_0$ is true:

$$P_{\theta_K}(Y \in \mathbf{Y}^c)$$

- Power is 1 minus the probability of type 2 error.

▶ **Power and size**

- Lowering size must lower power.
- You lower size by shrinking rejection region in the sense that the new one is contained in the old.
- A smaller region must mean lower probability of rejecting under any θ whether in Θ_0 or not.

▶ **Size vs. nominal size**

- In this class, *size* will always mean the definition above, which is also sometimes called exact size.
- The distinction is between size and size under some asymptotic approximation.
- We'll call the asymptotically justified notion *nominal size*

▶ **Pivotal test statistic**

- A test statistic is called pivotal if its distribution under the null hypothesis does not depend on any unknown parameters.

▶ **Consistent test**

- A consistent test is one for which the rejection probability converges to 1 with sample size for every fixed $\theta_K \notin \Theta_0$.

  However the null is false, you reject eventually using this test.

▶ **Confidence set/confidence interval**

- A confidence set procedure is a mapping from $\mathcal{Y}$ to subsets of Θ.

$$\psi : \mathcal{Y} \to S(\Theta)$$

  where $S(\Theta)$ is the collection of subsets of Θ.
- The procedure attains confidence level $100(1-\alpha)$ percent if:

$$P_\theta(\theta \in \psi(Y)) \geq 1 - \alpha \text{ for all } \theta \in \Theta$$

- That is, no matter which θ is true, $\psi(Y)$ contains or covers the true θ at least $100(1-\alpha)$ percent of the time.

▶ **Confidence set vs. interval**

- If θ is a scalar and the sets in questions are all intervals, then we say 'confidence interval' rather than the more general 'confidence set'

- In the interval case, we can think of the procedure as defined by two statistics, $c_l(Y)$ and $c_u(Y)$ giving the lower and upper bounds or values for the interval.
- For most of the rest of this note, I'll mainly refer to confidence intervals when the distinction is not essential

  just for clarity and simplicity.

▶ **Confidence intervals imply tests**

- Suppose we have a valid $100(1-\alpha)$ percent confidence interval for β, $[c_l(Y), c_u(Y)]$
- Define a test by, reject $H_0 : \beta = \beta_0$ whenever $\beta_0 \notin [c_l(Y), c_u(Y)]$
- This clearly is a size α test procedure.

  You should verify that you understand this.

▶ **Collection of Tests imply confidence intervals**

- DGP, $Y \sim P_\theta, \theta \in \Theta$.
- Parameter of interest is $\beta = g(\theta)$ and the parameter space for β is $\beta \in B = g(\Theta)$
- Suppose we have a test statistic $\phi(\beta_0, Y)$ for testing the point hypothesis $H_0 : \beta = \beta_0$

  And this test statistic is defined for every $\beta_0 \in B$.

- Pick a test size, e.g., 10 percent
- Run the test for $H_0 : \beta = \beta_0$ for every $\beta_0 \in B$.
- Collect the set of β_0s not rejected (nr):

$$B_{nr} = \{\beta_0 \in B | \text{test not reject } H_0 : \beta = \beta_0 \text{ at 10 pct. level}\}$$

- This will be a valid 90 percent confidence *set* (can't be sure its an interval)

▶ **Aside:: Why?**

- Take any $\tilde{\beta} \in B$ that is left out of this set (i.e., $H_0 : \beta = \tilde{\beta}$ is rejected).
- By definition of size of the test, when $\tilde{\beta}$ is true, this happens at most 10 percent of the time.

▶ **Question: What confidence interval does the standard (two-tailed) regression t test lead to?**

▶ **t-tests and confidence intervals**

- The t test of $H_0 : \beta = \beta_0$ rejects when (in what I hope is obvious notation):

$$\tau = |\frac{\hat{\beta} - \beta_0}{\hat{\sigma}}| > c$$

for the relevant c.

- As β_0 moves away from $\hat{\beta}$ in either direction, the statistic increases.
- In fact, it's easy to derive that the 'nonrejection region' will be in,

$$\hat{\beta} \pm c\hat{\sigma}$$

- This is probably the first confidence interval you learned.

  And the fact that $c = 1.96$ is the 95% 2-tailed critical value leads to the familiar 'plus or minus 2 standard errors. . . '

▶ **Confidence intervals and estimators**

▶ **Confidence intervals and estimators**

- If we take the confidence level up, the interval grows in length

  if you want to be more confident to cover the truth, you need a longer interval.

- And if you take the confidence level down to zero, this confidence interval collapses to a point

  That is, $\hat{\beta} \pm 0$.

- Not all confidence intervals are centered on some natural point estimator
- But lots of (most?) confidence interval procedures will collapse to a point as you take the confidence level to zero.
- And this point may be an interesting estimator that one might not have arrived at through other means.

▶ **Definition**

- Definition: A confidence interval is equi-tailed if it is equally likely to lie entirely below or lie entirely above the true value.
- Definition: An estimator is called median unbiased if it is equally likely to be above and below the true value.

  That is $P_\theta(\hat{\beta} > \beta) = 0.5$ for all θ where $\beta = g(\theta)$.

▶ **Median unbiased estimators and equi-tailed confidence intervals**

- Suppose we have an equi-tailed confidence interval procedure that is defined for all confidence levels
- And that converges to a single point as the confidence level is taken to zero.
- By definition, this point is a median unbiased estimate of the parameter.

    Even at confidence level zero, it's equi-tailed and thereby ...

▶ **Aside:: Median unbiased estimator of largest root**

- The leading case here in time series is in the case of unit root problems.
- We form a median unbiased estimator of the root nearest the unit circle.

    James H. Stock, Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, Journal of Monetary Economics, Volume 28, Issue 3, December 1991, Pages 435-459,

    go

    http://www.sciencedirect.com/science/article/pii/030439329190034L

- Also see: Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model, Stock and Watson, JASA, v93 n441 pp. 349-358

    go

    http://www.jstor.org/stable/2669631

- We'll get to the guts of these later.

▶ **Now, a simple intuition tester**

▶ **A simple intuition tester**

- Suppose you have an estimator, $\hat{\beta}$, and
    - Suppose it's distribution is skewed heavily to the right
    - And suppose that the true parameter, β^*, merely shifts the distribution of the estimator right or left but does not change this degree of skewness or other aspects of shape.
- On the sample at hand, the point estimate is, say, $\hat{\beta} = 7$

▶ **Question: Based on your knowledge of the distribution of the estimator, will the natural confidence interval reach further above or reach further below 7?**

▶ **Answer**

- It will reach further below 7

    In relating estimators to confidence intervals, the upper or right tail of the estimator distribution 'determines' the left or lower boudary of the confidence interval.

- That is, when the distribution of the estimator is skewed to the right, the natural confidence interval reaches farther to the left of the estimator.

▶ **Huh?!?**

- The distribution of the estimator is skewed to the right.
- This means that $\hat{\beta}$ tends to be much bigger than the true value.
- This means that you should look for the true β to be smaller than (to the left of, on the number line) the estimator.
- We will see later that some sort of intuitive bootstrap techniques effectively get this wrong.

▶ **Finally: asymptotics**

- So far we have used only exact language and completely ignored the fact that our inference in macro will generally be based on asymptotic approximations.
- We turn now to the asymptotic context

▶ **Asymptotically pivotal**

- A statistic is asymptotically pivotal if it converges in distribution to a random variable with a distribution that depends on no unknown parameters.
- That is, for the test statistic $\phi_T(Y)$:

$$\phi_T(Y) \to_d \phi$$

  and the distribution of ϕ involves no unknown parameters.
- If the distribution of ϕ_T is $F_{\theta,T}$ and that of ϕ is called ϕ_∞ (no θ), we might write,

$$F_{\theta,T} \Rightarrow F_\infty$$

  or

$$F_{\theta,T}(c) \to F_\infty(c)$$

  at every c where F_∞ is continuous, and $\to$ means standard calculus limit.
- Of course, a main purpose for exploring asymptotics is that many test statistics are asymptotically pivotal allowing us to make nominal size statements.

▶ **Nominal size**

- When our test statistics is not pivotal, computing exact size is a pain because the rejection probability may differ in unknown ways as we move around the null (choose different $\theta \in \Theta_0$).

- Thus, it may be impractical to evaluate the $\sup_{\theta \in \Theta_0} P_\theta$(reject null).
- If the statistic is asymptotically pivotal, then under the asymptotic distribution the probability of rejecting the null does not vary with $\theta \in \Theta_0$.
- The probability of rejecting the null under the asymptotic distribution is called the nominal size of the test.

▶ **Aside:: Usage**

- In practice, many folks drop the 'nominal' since we almost never have exact size.
- Don't do this.

  Dropping the 'nominal' fosters the tendency to forget we are using approximations and need to check their accuracy.

▶ **Asymp. pivotal and the CAN framework**

- In the CAN framework, our estimators are asymptotically normal, but are not asymp. pivotal.
- Their asym. dist. depends on the asymptotic variance-covariance matrix, which is unknown.

  Thus, $\sqrt{T}(\hat{\theta} - \theta)$ is not asymp. pivotal
- If we have a consistent estimator of the asymptotic variance-covariance matrix, however, we can form a pivotal statistic by dividing the point estimate by its estimated standard error

  Or dividing the squared estimate by its estimated variance.
- This sort of logic makes our standard asymptotically piovatal t, χ^2, F distributed test statistics
- The key is that, because the variance-covariance estimate is consistent, the variability of that estimate won't affect the asymptotic distribution of the ratio.
- This is why, once again, the CAN framework really needs to be called CAN-WCEAVCM

▶ **Power in the asymptotic context**

▶ **Power in the asymptotic context**

- To choose among different tests, we generally seek the one with the greatest power for a given size.

  constrained optimization: minimize type II error subject to a fixed type I error rate.
- Once we are in the asymptotic context, however, choosing becomes a bit more subtle
- Most standard tests are consistent.

  in a large enough sample, any decent test gets the right answer.

- If they are all right asymptotically, we need a sharper criterion to choose among asymptotically justified tests.
- Various more refined comparisons of power have evolved
- The most common is considering power against local alternatives.

  The alternative in question gets closer to truth as the sample size grows: as you get more information, focus on harder questions...

► **Aside:: Remember, in the CAN framework**

- Remember that the key to the can framework is starting with a consistent estimator.
- And we assume that our sample size is large enough for such results to matter.
- In short, we presume we are starting with what is nearly the right answer.
- About all that is left is to distinguish among answers in a small neighborhood of truth.

► **Pitman drift/local alternative**

- Suppose truth is $\beta = \beta^*$.
- We are looking to reject wrong values such as $H_K : \beta = \beta_k \neq \beta^*$
- Consider power against a hypothesis that evolves with the sample size:

$$H_{k,T} : \beta = \beta_{k,T} \equiv \beta^* + k/\sqrt{T}$$

  In particular, as the sample size grows, this alternative gets closer to the null.

- In a natural language: 'the alt. remains in a root-T neighborhood of β^*'
- This is called Pitman drift in the hypothesis

  Named after E. Pitman who may have first used this device (unless, e.g., Neyman did).

► **Local alternatives in the CAN framework**

- Our CAN estimator, $\hat{\beta}$, satisfies

$$\sqrt{T}(\hat{\beta} - \beta^*) \overset{a}{\sim} N(0, V)$$

- Consider the distribution of $\sqrt{T}(\hat{\beta} - \beta_k)$ for a fixed β_k
- $$\sqrt{T}(\hat{\beta} - \beta_k) = \sqrt{T}(\hat{\beta} - \beta^*) + \sqrt{T}(\beta^* - \beta_k)$$
- So that first term is asymptotically normal, and that second term is deterministically diverging to ($\pm$) infinity at the rate of $\sqrt{T}$

- It is this divergence that makes the standard test consistent against any alternative no matter how close to β.

▶ **Under hypoth. with Pitman drift**

- The location of the statisitc under $\beta_{k,T} = \beta^* + k/\sqrt{T}$ will be governed by

$$\begin{aligned} \sqrt{T}(\beta(k,T) - \beta^*) &= \sqrt{T}(k/\sqrt{T}) \\ &= k \end{aligned}$$

  which is to say, the stat. will not diverge.

- Thus, the tests won't be consistent against local alternatives

  (this should make sense to you).

- Thus, we can pick a local alternative of this sort and ask which test has the most power against local alternatives.
- This leads to the notion of locally optimal tests

▶ **Optimality and LR tests asymptotically...**

▶ **Optimality and LR tests asymptotically...**

- We know that likelihood based estimators and tests are often best in some sense.
- Once we leave the world of point hypothesis, we might want to makes statements about being best against all alternatives
- The LR test is most powerful or uniformly most powerful in certain well-known contexts.
- But, the assumptions under which there exist exact UMP (uniformly most powerful) tests are very strict
- Generally the distribution of the test statistic under the alternative depends on the particulars of the alternative

  the test that has best power in one direction is not best in some other direction.

- But asymptotically some of these issues may diminish.
- In a local neighborhood of truth, it may be true that the the stuff that keeps there from being one best test has become unimportant.
- That is, there may be a single best test in a $\sqrt{T}$ neighborhood of truth.

  And in a good world, the LR test would have this property if anything does.

- Well, we *almost* live in a good world.

We need to limit the universe of tests a bit to make the statement

- The LR test is often the locally most powerful invariant test.

  See Engle's lovely Handbook chapter 13 on the trinity of tests go

  http://dx.doi.org/10.1016/S1573-4412(84)02005-5

▶ **Aside:: Invariant, unbiased, similar, etc.**

- As we get deeper into general cases, it is often impossible to prove results across the universe of tests
- And we come up with technical bases for limiting the class of tests, thereby allowing us to prove results.
- Conditions that often arise are limitations to 'unbiased tests', 'invariant tests', 'similar tests'
- In this class, I'm not going to get into these issues and these words won't come up much.

  Unless and until you learn these concepts, just think of them as limiting the class of tests to those that are well-behaved in some technical sense.

- As with any vague notion of 'well-behaved' there are often deep questions about whether or not it makes *practical sense* to limit the discussion to things that are 'well-behaved' in that technical sense.

  And these questions arise regarding all the concepts I've just mentioned.

- The best source on these topics, is Lehmann, 'Testing Statistical Hypotheses'

▶ **Finally, Other ways to distinguish asymptotically among tests**

- There are other ways to distinguish among consistent tests.

  I'll just mention one concept for completeness.

- Bahadur defined a notion called the slope of a test.

  Interesting guy Raghu Bahdur, google him.

- In applied work, you seldom hear this term
- Geweke drew the topic more into the practical realm by defining a related concept the approximate slope.

  The Approximate Slopes of Econometric Tests John Geweke Econometrica Vol. 49, No. 6 (Nov., 1981), pp. 1427-1442 go

  http://www.jstor.org/stable/1911409

- This is just a teaser about things you might want to dig into on the deeper structure of testing

▶ **Wrapping up**

- This was a whirlwind tour across a lot of conceptual issues relating tests, confidence intervals and estimators.
- I am presuming that much of this is familiar to you.
- And much of the stuff that wasn't familiar, I just previewed for later in this class

  or later in your more general studies.