Problem set 7
ANSWERS
607: Applied Macroeconometrics
Fall 2016
Jon Faust

The following is due at the beginning of next class. You can turn in any paper in my mailbox or in class; email me and requested computer work. You may work in groups; hand in a single submission for the group. The submission should list those who contributed.

Note 1: This is a new (well, a greatly changed) problem set. Please help me proofread it and let me know if there are confusing or ridiculous bits.

Note 2: This problem raises breaks tests, which we have not talked about and is related to the issue of testing for breaks in correlation, which we started on last time. Two papers you will want to look are these:

Bruce Hansen's paper introduces the concepts in this area.

My paper with Brian Doyle on breaks in the correlation of GDP growth across countries is an application of what I think of as fairly careful econometrics. Well-motivated problem, careful treatment of several subtle econometric problems, bootstrap inference, and a Monte Carlo to check whether the inference approach is plausibly reliable.

Bruce E. Hansen; The New Econometrics of Structural Change: Dating Breaks in U.S. Labour Productivity; JOURNAL OF ECONOMIC PERSPECTIVES; VOL. 15, NO. 4, FALL 2001; (pp. 117-128)

https://www.aeaweb.org/articles?id=10.1257/jep.15.4.117

Doyle, B., and Faust J. 2005. Breaks in the variability and co-movement of G-7 economic growth; Review of Economics and Statistics, 7(4), Nov. 721-740.

https://www.jstor.org/stable/40042889

1. Terms

   (a) Chow test for structural break

Answer/comment

Partition the sample into two parts. The maintained model is that the parameters are constant in both subsamples. The null hypothesis adds that they are equal across the subsamples. The Chow test is the natural Wald test of this hypotheses: freely estimate the coefficients in the two samples, giving $\hat{\theta}_1$ and $\hat{\theta}_2$. The statistic is then some version of

$$W = (\hat{\theta}_1 - \hat{\theta}_2)\hat{V}^{-1}(\hat{\theta}_1 - \hat{\theta}_2)$$

and $\hat{V}$ is a consistent estimate of the variance of the difference. As described in the introductory part of the course, this is a generalized distance measure of the distance between the unrestricted estimates of the parameter values in the two subsamples. The bigger this distance, the less likely the null is true.

This answer is very abstract because I trust you could all look up the formula. I won't bother to type it myself. I want to emphasize a substantive way of thinking about what the statistic is.

Like all standard Chow tests, we can write an F or a χ^2 form.

(b) sup $-F$-style Chow test for structural break at unknown date

Answer/comment

Pick some t_{min} and t_{max} that are the earliest and latest possible break dates in the sample. This must each allow the model to be estimated on the subsamples on each side of the break (we need more observations than parameters).

Compute the F-form of the Chow test for each break date between t_{min} and t_{max}. Your test Take as your overall test statistic, the maximum of these. The distribution of this statistic will be nonstandard (none of the distributions that commonly arise in the CAN framework) under the null.

---

2. Structural breaks. You have a regression model:

$$y_t = x_t'\beta + \varepsilon_t$$

You want to test for a structural break in β occurring between observations t_b and $t_b + 1$.

(a) Calling the values in the two subsamples β_1 and β_2, write a single regression such that the test of $H_0 : \beta_1 = \beta_2$ is a simple coefficient restriction on the coefficients of a single regression.

Answer/comment

This empahsizes an earlier point: in the linear model world everything can be written as a single equation (breaks, simulataneous systems, etc.)

$$Y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon$$

where X_1 is the matrix of X values in the first subsample and X_2 the values in the second.

(b) What is the natrual Wald test of H_0? How is it distirbuted? What is its relation to the F-test described in your sup $-F$ answer.

Answer/comment

Calling

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

we have that

$$\sqrt{T}(\hat{\beta} - \beta) = N(0, V)$$

for some V. And the hypothesis of no change in the βs can be written as

$$R\beta = 0$$

Thus, the natural Wald test is

$$W = T(\hat{\beta} - \beta)' R' \hat{V}^{-1} R(\hat{\beta} - \beta)$$

where $\hat{V}$ is consistent for V. The statistic is asymptotically $\chi^2_{(k)}$ under the null, where k is the number of linearly independent rows in R, which in this case means k is the number of regressors, X. Every Wald test can be put in a χ^2 or F form. Intuitively, the χ^2 form treats $\hat{V}$ as fixed at the true value. Because $\hat{V}$ is consistent, this treatment is fine asymptotically. The F form takes into account variability in $\hat{V}$.

In particular, the F-form is,

$$W_F = \frac{(SSR - SSU)/k}{SSU/(T-k)}$$

where SSR and SSU are the restricted and unrestricted sum of squared residuals, respectively, and k is the number of linearly indep. restrictions under the null. You should be able to show that this is a monotonic transform of the Wald statistic.

(c) You don't know the break date, t_b. You look at the data and see a likely break date and run a nominal 5 percent Chow test for a break at that date. Why will the test of this procedure tend do have size greater than 5 percent? (Note: View the procedure as: look at the data, pick what looks to be a breakdate, then run the test using that break date.)

Answer/comment

Let's us suppose that the test at the fixed breakdate is properly distributed: that is, it would deliver 5 percent rejections under the null.

The procedure just described will tend to deliver a weakly larger value for the test statistic under the null of no break. Under the null, there is no reason to suppose that the break at the fixed date delivers the highest value of the test statistic. The highest value will occur (under the null) at different places depending on sampling fluctuation: the break will be placed where it helps most in soaking up sampling fluctuation in this sample.

In the limit, suppose that our peeking at the data always delivered the maximum possible value of the test statistic on this sample over all possible break dates. In this case, we know that the peeking procedure generates a weakly bigger value for the test statistic.

If the value is weakly greater (and strictly greater on most samples), the null must be rejected more often than under the fixed break.

(d) How does the procedure in the sup $-F$ test compare to the procedure in the previous part?

Answer/comment

Obviously, it just formalizes the 'peek and put the break where it seems most needed' procedures.

The sup F analysis shows an important lesson. When we have a complicated, data-based procedure, we can sometimes formalize that procedure allowing us to analyze its properties. In the 1960s and 1970s we only ran break tests when we thought we saw breaks in the data. We tended to find too many breaks (that is, too many false positives of break).

Formalizing the procedure as picking the $sup - F$ allowed us to pick a critical value that did better in delivering appropriate size.

(e) Breaks in a slightly different environoment. Now the model is:

$$y_t = x_t'\beta + z_t'\gamma + \varepsilon_t$$

The maintained assumption is that there is definitely one and only one break. And in particular, at least one element of β changes at that date. The parameter γ may or may not have a break.

You use some sensible approach for picking a likely break date. For example, you run the sup $-F$ test allowing for breaks in any of the βs or γ and take the date of the largest test statistic as t_b.

Then you use that breakdate for the date in a standard nominal 5 percent Chow test for constancy of γ only (in the notation of the previous problem: $H_0 : \gamma_1 = \gamma_2$).

Make an argument that this test of the γ will have proper size asymptotically.

The formal reasoning here is sketched in the Doyle-Faust paper and the papers cited therein.

The basic intuition is that if you know that there is a break, then as the sample size gets large you will find the right location of the break (in a particular sense) with high probability. Here we have to be a bit careful. You say that the break occurs a fraction r of the way through the sample, say 40 percent of the way through. Letting the sample size grow, you'll tend to get the share r correct based on the break in β alone.

Having nailed down the location in the break based on the βs, we don't have that freedom to soak up sampling fluctuation in the zs by moving the break around. Thus, you are effectively running a Chow test at a fixed date. Perhaps a more informative way of saying it, you are NOT running the test at a breakdate chosen to account for maximum sampling fluctuation in the zs.

---

3. Classroom presentation. This continues from our sample correlation Monte Carlo in ps6 (presentation should cover both).

   Add the three standard bootstrap confidence intervals to the Monte Carlo from last time.

   The three confidence interval procedures are the percentile, other percentile, and percentile-t. The bootstrap DGP is the XY-resampling scheme (resample rows of the data matrix with replacement).

   Note. Creating one draw of N observtions from a $(T \times 2)$ data matrix can be done very easily:

```
T=100;
N=100;
XY = randn(T,2);      % fake data for illustration

newXY = XY[ ceil(T*rand(N,1), :];  % the new sample.
```

   This relies on the fact that if A is a matrix in Matlab, $A[[25]', :]$ is a matrix composed of the second and fifth row of A.

   For various sample sizes and values of the population correlation, and for each of the 3 confidence interval procedures, report the Monte Carlo estimate of the coverage, and the share of draws in which the population value lies to the left and to the right of the confidence interval. These three numbers obviously add to 1. Also include a fourth confidence interval procedure: the confidence interval based on the assumption that the t statistic is distributed $N(0, 1)$—that is, the percentile-t confidence interval under conventional asymptotics.

   Answer/comment
   Here is what you should have thought ex ante: From last time, we figure that we'll find that the conventional asymptotic approach is pretty bad both in overall coverage and in getting the two tail probabilities to be equal.

The percentile confidence interval doesn't studentize and thereby may be somewhat better, but its a bit hard to intuit how much better.

The other percentile confidence interval? Well, that one is backwards and the statistic has an asymmetric distribution so it should be worse than the percentile and who knows relative to the percentile-t.

The results will be in the classroom presentation.

Note: As I said in lecture, the other percentile bootstrap is backwards and outside the symmetric case makes little sense. There are however various fixups that fix this problem. One of those is bootstrap iteration. Use a second bootstrap to overcome shortcomings in the original bootstrap. There is some good theory expalining why this works.

As it turns out, the iterated other percentile bootstrap has been found to perform well on correlations.

The Doyle-Faust paper looks at both percentile-t and iterated other percentile confidence intervals.
