607
Breaks

Jon Faust

http://e105.org/e607

October 24, 2016

- **Readings**
  - Wright notes, pp88-92
  - And articles cited throughout.
- **Note:**
  - We will use this lecture as a way to introducce by hand waving the notion of functional central limit theorems.
  - Remember that exploiting higher order terms in asymptotic expansions and FCLTs are the two main new tools we take on board in going from advance econometrics to modern advanced econometrics.
- **This lecture**
  - Presents a family of linear-with-breaks models
  - Start with linear equation:
    $$y_t = x_t'\beta + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2)$$
    $t = 1, \ldots, T$
  - Now add the possibility that the slope coefficients change at a known breakpoint, T_b:
    $$\begin{aligned} y_t &= x_t'\beta_1 + \varepsilon_t \quad t \le T_b \\ y_t &= x_t'\beta_2 + \varepsilon_t \quad t > T_b \end{aligned}$$
    $\varepsilon_t \sim iidN(0, \sigma^2)$

- Or write it this way:

$$y_t = x_t'\beta(t) + \varepsilon_t$$

  where $\beta(t) = \beta_1$, $t \leq T_b$ and $\beta(t) = \beta_2$ otherwise.

▶ **When would this model be appropriate?**

- Might use this when there was an important policy change at a known point

  Exchange rates before and after breakdown of Bretton Woods or launch of euro

▶ **Aside:: other wrinkles**

- Could allow σ^2 to change as well, but will leave that fixed for notational simplicity
- Could also allow more than one break, with all breaks happening at known times

  Stick with one break in only β for now

▶ **No new analytics needed**

- We can the model as

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

- That is

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

- The variance-covariance matrix for the resulting εs will have some special features, but we already know how to deal with general vcov matrices.

  GLS, robust std. errs. etc.

- Emphasize: the break adds nothing to the estimation and testing issues we have already considered

▶ **Breaks testing**

- The classic way to test whether $\beta_1 = \beta_2$ is called a Chow test.
- The Wald form of the Chow test is just the natuaral Wald test for testing $H_0 : \beta_1 = \beta_2$ in the stacked regression:

$$W = T(R\hat{\beta} - 0)'(R'\hat{\Sigma}R)^{-1}(R\hat{\beta} - 0)$$

  where $\hat{\beta}$ is OLS est. of $\tilde{\beta}$, $\hat{\Sigma}$ the estimated vcov matrix, and R is?

  $R = [I : -I]$ (I $k \times k$ identity matrix if β_1 and β_2 $k \times 1$)

- Under the null (plus standard assumptions) W will be asymptotically distributed $\chi^2_{(k)}$

- **Aside:: Other tests**
  - The F form of the Wald test is often used.
  - Both of these are asymptotically equivalent to the natural LR test in this case.
  - As stated so far, we didn't allow for a break in the variance of ε as well, but adding that adds no major complication.
- **In short**
  - No new issues in estimation and testing are raised by the generalization to a break at a fixed point.
  - Let's add some richness
- **Single break, unknown time**
  - The single break occurs at an unknown breakpoint
  - Emphasize: we are not viewing the breakpoint as stochastic, simply as unknown to the econometrician.

    Just as we treat β as fixed but unknown in the classical (but not Bayesian) world
  - The model is same as before, except now we have to estimate $T_b \in \{1, \ldots, T\}$

    That is, we have one more parameter
- **How would you estimate this?**
  - Estimate just as we'd estimate anything: ML, EGMM, etc.
  - We want $\max_\theta L(X; \theta)$ where $\theta = \{\beta, \sigma^2, T_b\}$
  - Note: we know how to find $\max_\theta L(X; \theta | T_b = t)$

    we are back in the fixed break model.
  - We can run each conditional model, and then pick the best among these.
  - Thus, we can run the OLS regression for all possible break dates, saving $L_{T_b=t}$, for each possible date and your ML estimate is the one with the highest likelihood
- **Unlike the fixed breakpoint case, in the unknown breakpoint case, a few special issues arise when we move from point estimation to testing.**
- **two issues in testing**
  - In testing a classic problem arises: we have a restricted and an unrestricted model, but we have a parameter that is not identified (or is simply undefined) under the null.

- In this case, under the null, there is no break so there is no true value of the break date.
- Whenever it arises that some parameter is identified under the alterantive but not the null, standard CAN framework-style testing results break down.

▶ **Intuition**

- The Chow test we described above was a test that $\beta_1 = \beta_2$ where β_1 prevails from $t = 1, \ldots, T_b$ and β_2 prevails thereafter.
- Under the null of no break, we don't even know where to put T_b in order to compute the test statistic.
- Suppose instead we are running an LR test; thus, we maximized the likelihood both under the restricted (no break) and unrestricted models.
- In standard cases, even when the null is true, all the parameters of the unrestricted model are CAN

  they converge to something and when normalized are asymptotically normal.

- But what breakdate will be found under the unrestricted model when the null is true?

  It can't converge to anything. In any given sample, there will be some breakdate chosen, but there is not 'true' value nailing it down.

- These are all intuitive attempts to convince you of the analytic result that standard distributional results won't hold in this case.
- Arises regularly in many contexts
- Elegant treatment Andrews, Ploberger:

  Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative Econometrica, v62 n6 (1994), pp. 1383-1414 go

  http://www.jstor.org/stable/2951753

- Applied to breaks in specif.:

  Andrews, Tests for Parameter Instability and Structural Change With Unknown Change Point Econometrica, v61, n4, pp. 821-856 go

  http://www.jstor.org/stable/2951764

- A very nice less technical exposition by Bruce Hansen.

  The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity J Ec. Persp., v15, n4, 117-128 go

  http://www.jstor.org/stable/2696520

- **A practical problem**
  - At a practical level, what traditionally happened was that folks...
    - Looks at the data or preliminary results
    - Saw what looked like a break, and then
    - Ran a standard Chow test for a break at that date
  - This is essentially a way of maximizing pre-test bias.

    you only run the test in cases where you first see ssomething that looks like a break at the point in practice. A good way to maximize type I errors.

- **An idea**
  - In an important sense, the problem here is that we have not defined how the break is determined under the null.
  - So long as we specify that determination in some systematic manner we can hope to do statistics.

- **Specifically,**
  - Take the vector of test statisitics for each possible break date.

    more specifically, for each possible breakdate where it is possible to compute the statistic (We need enough obs. in each sample to estimate the model so the break can't be too near either endpoint of the sample.)
  - Call this vector of, say, Wald statistics $W = (W_f, \ldots, W_\ell)$

    where f and ℓ are the index of the first and last feasible break dates.
  - Specify a systematic manner to map this vector into a scalar:

$$W^* = \phi(W)$$

  - Now we have a hope of doing some rather conventional statistical analysis

    conventional at least from the standpoint of modern advanced econometrics.

- **Of course we need asymptotics**
  - As usual, we need some asymptotic theory to guide us.
  - It turns out that we can use something called a functional central limit theorem (FCLT).
  - FCLT's will be very important in our discussion of modern advanced econometrics

    And, in particular, in our discussion of unit roots.

▶ **Aside:: Standard Brownian motion**

- A standard brownian motion is a continuous time process on $r = [0, 1]$, $W(r)$
- Any increment $W(r) - W(l)$, $r > l$ is $N(0, (r - l))$
- Nonoverlapping increments are jointly normal and uncorrelated.
- $W(0) = 0$.

▶ **Aside:: Brownian motion: analogy to discrete time random walk**

- Discrete time, standard Gaussian r.w.:

$$x_t = \sum_{s=1}^{t} \varepsilon_s, \quad \varepsilon \sim iidN(0, 1)$$

- $y_{t_2} - y_{t_1}$—the increment between t_1 and t_2—is just the sum of the εs occuring in the interval.
- Thus, the variance is just the length of the interval, $(t_1 - t_2)$
- And nonoverlapping increments are jointly normal and uncorrelated.
- Thus, the Brownian motion looks like a continuous time Gaussian random walk.
- $B(0) = 0$

▶ **FCLT: in words**

- Suppose we have a discrete set of random variables, $\Xi = (\xi_1, \ldots, \xi_T)'$.
- And it is coherent to think of the number of elements in this vector going to infinity

  so that we can do asymptotics

- Functional central limit theorems give condiions under which

$$\tilde{\Xi}(r) \to_d W(r), r \in [0, 1]$$

  where $\tilde{\Xi}(r) = \xi_{[Tr]}$ and [] means integer part of.

- And a continuous mapping theorem then implies that statistics $\phi(\Xi)$ are distributed as $\phi(W(r))$.

▶ **FCLTs: the big idea**

- We'll talk much more about this.
- For now just accept that we can sometimes map a set of statistics indexed by $1, \ldots, T$ into a function indexed on $[0, 1]$.
- And as T gets large, a kind of central limit theorem allows us to characterize the behavior of that whole function.

And hence by a continuous mapping theorem, statistics that can be writted as a function of that random function.

- **FCLTs and breaks at unknown times**

  - We define $\tilde{W}(r)$ to be the function on [01] defined by

$$\tilde{W}(r) = W_{[Tr]}$$

where the Ws are our Wald statistics, r is interpreted as a break happening the share r of the way through the sample, and $[x]$ means the integer part of x.

  - And through some magic and applicaion of an FCLT, we find that,

$$\tilde{W}(r) \rightarrow_d FBM$$

function of Brownian motions.

  - Thus, under our continuous mapping theorem,

$$\phi(\tilde{W}(r)) \rightarrow_d \phi(FBM)$$

- **Interesing ϕs**

  - Since the data won't sensibly pick among the different possible test statistics, we have to choose.

That is, we pick ϕ.

  - What are some good choices.

- $\sup W$

  - One natural choice is to simply take our statistic to be the biggest of the individual Ws

I say biggest presuming that we are using a stat. such as the Wald stat. that rejects for large values.

  - Notice that this is arguably what the profession was doing a sloppy job of doing in the past.

We tended to place the break where the data made it look like a break had occurred.

  - In an important sense, the critical values for this test will be the ones that appropriately adjust for pre-test bias or data snooping bias.

- **Example from the Hansen article**

  - Hansen's JEP article provides a nice intuitive version

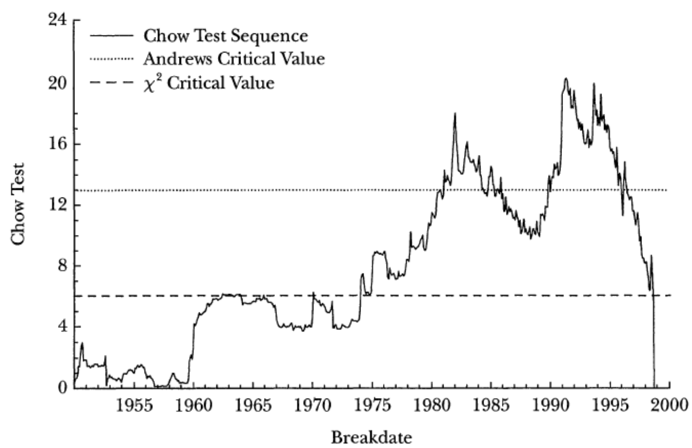
The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity J Ec. Persp., v15, n4, 117-128 go

http://www.jstor.org/stable/2696520

- **Example from the good old days**
  - After the first oil shock in the mid-70s we got the next 10 or so years of data
  - Saw that productivity seemed to have fallen

    And our Chow tests "confirmed" this
- **Hansen JEP article**
  - In example, Hansen estimates an AR(1) with break in growth of labor productivity
- **Result**

Figure 1
Testing for Structural Change of Unknown Timing: Chow Test Sequence as a Function of Breakdate



- **Thus,**
  - For many years, we were claiming we had strong evidence of a mid-1970s break when in fact the evidence was very weak
  - Now the best techniques probably suggest there was a break sometime in the 1970s
- **Comment on today: secular stagnation?**
  - Maybe, but we won't have real evidence of it for many many years.
  - Note that this discussion is all intimately tied up with the unit root question

A break is a persistent change in the behavior of a series.

- It only seems different when we limit our focus to a single change
- But that single change may be a useful approximation for use, but there is no reason to believe there are such changes in reality.
- Given expectations, reaction times, adjustment costs and learning, there are strong reasons to believe that there are no such sharp changes.

useful approximation? maybe. Truth. Nope. But truth is over rated in this area.

▶ A deeper point in the logic of testing

- This topic allows us to introduce one more fundamental idea in testing theory.
- Suppose we have a bunch of test statistics
- e.g., Break tests for different break dates
- This is a verion of a more general phenomenon where we can create different tests under the same null directed at different verions of the alternative
- For example, in many cases with a composite alternative and point null we could create a different Neyman-Person-style point optimal test for each different version of the alternative.

(Only in the rare instances when there is a uniformly most powerful test would this not give rise to a choice problem among possible tests)

▶ Optimizing

- If we had a well-defined substantive decision with a loss function, we could derive an unambiguously optimal combination for the context at hand.
- But if we had a well-defined decision and loss function, we probably should have started from that in the first place to choose an inference approach
- In macro research, our goal is something like 'learning about how the economy works' (or, let's face it, getting tenure)
- These goals are a bit too loosely specified to nail down an optimal combination.

▶ More ad hoc, but appealing solutions

- We still can think about some sort of more *ad hoc* optimality criterion
- For example, we might consider maximizing weighted average power across all the versions of the alternative hypothesis

Where the weights reflect, uhhh, well something sensible.

- Andrews and Ploberger discuss this more fully.

- The version of optimal unit root testing originated by Elliot-Rothenberg-Stock is based in the analogous idea.

  Under local-to-unity asymptotics, standard tests are not pivotal, so there is no on test that makes sense for all c in $\rho = 1 - c/T$. But we can try to optimize some verion of weighted average power.

- This is a very powerful idea for generating attractive tests when there are multiple standard tests vying for our attention.

▶ **Finally a detail**

- As we noted above, in a finite sample, we cannot generally compute standard test statistic for a break, say, after the first observation.

  We don't have enough observations in the first sub-sample to estimate the model.

- This issue manifests itself even asymptotically.
- Remember, we reconceptualize the test as allowing a break at a share $r \in [0, 1]$ of the way through the sample.
- No matter how large is T, there will always be rs so near 0 or 1 that this problem manifests itself.
- Thus, we can do the asymptotic theory allowing r on $[0, 1]$, but it says what it must: the test is ill-behaved
- In practice, we have to consider $r \in [\kappa, 1 - \kappa]$, where κ is something like 0.1 or 0.15

  for $\kappa = 0.15$, the break doesn't occur in the first or last 15 percent of the sample.

- And the asymptotic distribution depends on κ—as κ moves to zero the null distribution moves to the right smoothly toward the ill-behaved case.
- Thus, we have a trade-off between controlling size when the break is near an endpoint vs. power.

▶ **Elaborations**

- Question: Is there a common break in multiple series.

  To test: Bai Lumsdaine Stock (1997) Bai (2000, Annals of Ec and Fi)

- We are talking about breaks in coefs of

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

- We have implicitly assumed that marginal process for the Xs doesn't break.

  If our relation for Y is stable, but the process for the Xs has changed, this can affect the dist. of the test for the stability of β.

- Thus, is there 1 break, timing unknown, Xs may have breaks

    Hansen (2000) gives the test and a bootstrap approach

- Are there multiple breaks, timing unknown

    Bai and Perron 1998, they give gauss code, differs by number of possible breaks

- And so forth

▶ **Summary**

- All of these elaborations are fairly straightforward once you understand the basics.

▶ **A different approach: Nyblom-style tests**

- Nyblom derived a test with certain local asymptotic optimality properties
- Bruce Hansen extended

    Nyblom, JASA, 89; Hansen, JBES, JPM, 92

▶ **Nyblom basics**

- OLS normal equations say that for each regressor i, that

$$\sum x_{ti}\hat{\varepsilon}_t \equiv \sum f_{ti} = 0$$

  where $f_{ti} \equiv x_{ti}\hat{\varepsilon}_t$

- Define partial sums of the fs:

$$S_{ti} = \sum_{s=1}^{t} f_{si}$$

    note $S_{Ti} = 0$ by definition

- $S_{Ti} = 0$, but for $T < 1$, under the null of no change (β_i is constant), S_{ti} should wander around zero as t changes.
- Picture
- Can base a test on how variable S_{ti} is around 0. The test
- The statistic:

$$\begin{aligned} \gamma &= T^{-1}\frac{\sum_{t=1}^{T} S_{it}^2}{V_i} \\ V_i &= \sum_{t=1}^{T} f_{it}^2 \end{aligned}$$

- If S wanders far from zero, γ will be large. Distribution under null
- The distribution of γ under the null is nonstandard, critical values in Hansen's two 1992 papers (JPM and JBES)

  JPM is stationary case, JBES allows some nonstationarities

- Can also test for stability of the variance (σ_ε^2)

  In thie case, $f_t = \hat{\varepsilon}_t^2 - s^2$ where s^2 is sample variance of $\hat{\varepsilon}$.

- Can also jointly test stability of any set of coefficients and σ_ε^2

  Including testing all βs and σ_ε^2 are jointly constant