607
Persistence: intro.

Jon Faust

http://e105.org/e607

November 30, 2015

- **Some history**
  - The profession spent the 30 years or so from the 50s through the 70s building up a lot of econometric results

    This was the first blossoming of macro modelling made possible by advances in thinking, computing power, and the emergence of good macro data.
  - This edifice basically collapsed around 1980 with theoretical critiques of Sims and Lucas, more practical critiques of Hendry and Leamer.
  - Playing a lesser role in the collapse was the realization that many results were not reliable because they ignored the persistence problem.
  - We'll be clearer about the what we mean by persistence problem later, but for now just think of it as an umbrella term for the fact that many CAN framework approximations are quite poor in the presence of data with high levels of serial correlation.
  - Much of the 1980s and 1990s was spent with economists and macro modellers sorting out the theory of building better macro models

    (Giving rise to DSGE models and to a focus on Euler equations, and, hence, GMM.)
  - Meanwhile, time series econometricians tried to sort out the persistence problem in time series
- **As with any upheaval in thought**
  - Lots of misguided stuff was stated as this complicated area was sorted out.
- **Two titles that sum it up**
  - Campbell and Perron: Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots

NBER Macro Annual, 1991 go

http://www.nber.org/chapters/c10983.pdf

- Christiano and Eichenbaum: Unit roots in GNP: Do we know, and do we care?

  Carnegie-Rochester Conference Series on Public Policy 02/1990; 32(1):7-61. DOI: 10.1016/0167-2231(90)90021-C

- The first of these essentially said that unit roots can be a pain from the standpoint of statistical inference, but give us some very powerful tools for choosing between economic theories.
- The second essentially says that the first should be Pitfalls and Pitfalls Masquerading as Opportunities.
- This is all very recent from the standpoint of scientific progress.
- Much research still is a bit confused on these points.
- Thus, you'll need some history of thought to help you understand and relate to the literature.

▶ **Phase 1 was understanding the asymptotics of processes with unit AR roots.**

▶ **Integrated processes**

- The discussion begins with the idea of order of integration.
- We say a variable is integrated of order d, $I(d)$, if after differencing the variable $d-1$ it is nonstationary and after differencing d times it is stationary.
- The standard random walk is $I(1)$.
- If inflation is stationary, then the price level is $I(1)$.
- If inflation has a random walk component, then the price level is $I(2)$.
- In macro we seldom run into things beyond $I(2)$

▶ **Aside:: Comment, I(1) vs. nonstationary**

- You all know the definition of stationarity (weak and strict) and not all violations involve integrated variables.
- Often we use the language stationary vs. nonstationary when the particular kind of nonstationarity we are referring to has random walk-ish or explosive—that is, an I(1)—character.
- Should pay close attention when nonstationary is used…

▶ **First key realization: the asymptotic distribution of many statistics changes discontinuously as we move from AR roots near 1 to roots of exactly 1; the exact unit root case gets quite complicated asymptotically**

▶ **Illustration:**

- Consider two Gaussian AR(1) processes that are independent but each have the same ρ parameter:

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \varepsilon_t \\
x_t &= \rho x_{t-1} + \nu_t
\end{aligned}
$$

ε and ν mutually uncorrelated, iid $N(0, \sigma^2)$.

▶ **Regress y on x**

- Suppose we run a regression

$$y_t = \beta x_t + u_t$$

- For all ρ, truth is $\beta = 0$.
- For $|\rho| < 1$,

$$
\begin{aligned}
\hat{\beta} &\to_p 0 \\
\sqrt{T}\hat{\beta} &\to_d \quad N(0, V) \text{for some } V \\
R^2 &\to_p \quad 0
\end{aligned}
$$

- For $\rho = 1$

$$
\begin{aligned}
\hat{\beta} &\to_d \quad FBM \\
\sqrt{T}\hat{\beta} &\to \quad \text{diverges} \\
R^2 &\to_d \quad FBM
\end{aligned}
$$

and the R^2 distribution is centered on about 0.2.

- FBM, as we'll explain, stands for a distribution that can be written as some function of Brownian motions.
- Summary
  - with $|\rho| < 1$, $\hat{\beta}$ and R^2 converge to 'true' values
  - with $\rho = 1$, neither $\hat{\beta}$ nor R^2 converge to single values at all. The distribution of $\sqrt{T}\hat{\beta}$ diverges.

▶ **Aside:: Spurious regression**

- The set of results associated with running one random walk on a separate independent random walk are known as the spurious regression case.

- Explored by Granger and Newbold, 1974

  Spurious Regressions in Econometrics, Journal of Econometrics, 2,111–120.

- Later formalized by Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, Journal of Econometrics, 33, 311-340.

▶ **Example 2**

- Run a regression of y_t on y_{t-1}, perhaps with a constant:

$$y_t = [\ \alpha + \ ]\ \ \beta y_{t-1} + u_t$$

- And define τ_ρ to be the standard t-test of $H_0 : \beta = \rho$.
- With $|\rho| < 1$, we have

$$\begin{aligned} \hat{\beta} &\to_p \rho \\ \sqrt{T}(\hat{\beta} - \rho) &\to_d N(0, V)\text{for some } V \\ \tau_\rho &\to_d N(0, 1) \end{aligned}$$

  and these results are invariant to whether or not there is a constant in the regression.

- With $\rho = 1$:

$$\begin{aligned} \hat{\beta} &\to_p \rho \\ \sqrt{T}(\hat{\beta} - \rho) &\to_p 0 \\ T(\hat{\beta} - \rho) &\to_d FBM \\ \tau_\rho &\to_d FBM \end{aligned}$$

  and the latter results differ (the functions of Brownian motions differ) depending on whether there is a constant in the regression.

- Thus, in this case $\hat{\beta}$ is super consistent

  Even blown up by $\sqrt{T}$, the distribution of $(\hat{\beta} - \rho)$ collapses

- The t statistic does not have its standard distribution and the distribution differs depending on whether the constant is estimated or is fixed at its true value of zero.

▶ **Huh?**

- We'll briefly review the analytics underlying these results in a separate note

- For now, I just want to emphasize that there is a discontinuous break in asymptotic distributions as we go from ρ arbitrarily close to 1 to $\rho = 1$.

▶ **A misguided binary view**

- Starting from the naive view that asymptotic results are always of practical relevance, a significant amount of work from 1980 through the early 1990s took a binary view
- In this view, every variable was I(1) or I(0), and the statistical approach we should take depends on which is the case.
- Thus, applied projects involved a diagnostic step at the start in which variables were labelled I(0) or I(1) and then analysts used whichever body of asymptotic results was relevant under that deterimination.

▶ **Nonsense**

- This approach was nonsense
- Return to the examples above.
- In any fixed sample size, as we move ρ smoothly to 1, the exact distribution of statistics changes continuously—the distribution shows no break at 1.
- Thought experiment: Fix $T = 100$, generate some εs and use the same εs to generate two AR(1) samples, one with $\rho = 1$, the other with ρ very close to 1.
- Obviously, as ρ approaches 1, these samples look alike.
- The distribution of statistics based on samples will vary continuously with ρ even in a neighborhood of $\rho = 1$.
- We can make a formal argument from continuity here

  Hint: think about the continuity of the likelihood for any sample of fixed size.

▶ **Thus,**

- We have one asymptotic theory valid for ρ arbitrarily close to 1
- And a different asymptotic theory for ρ exactly 1
- And there is some small neighborhood of 1 in which these two cannot both be good approximations

▶ **Cannot overemphasize: truth is over rated.**

▶ **Truth is over rated**

- I am omnicient and tell you some series is I(0).

- What good is this information if all you care about is distribution of statistics in sample sizes you care about?

  Answer: none. You still need know if the series is sufficiently persistent that statistics will behave as if the series is I(1)

- This goes on the list with lag length selection as an issue where truth is over rated.

  Remember: you have 50 observations and I tell you truth is an AR(30). Q: Of what value is this statement? A: No practical value.

▶ **So we throw out the dichotomous view as irrelevant to applied work. To do better it is useful to do some unit root unit root analytics**

▶ **Analytics**

- Let's compare some analytic results for a generic I(1) process y and and I(0) process x.

$$\begin{aligned} \tilde{A}(L)y_t &= B(L)\varepsilon_t \\ (1-L)A(L)y_t &= B(L)\varepsilon_t \\ (1-L)y_t &= C(L)\varepsilon_t \end{aligned}$$

  and

$$\begin{aligned} A(L)x_t &= B(L)\varepsilon_t \\ x_t &= C(L)\varepsilon_t \end{aligned}$$

  where $C(L) \equiv A(L)^{-1}B(L)$

- Where we are assuming that
  - $A(L)$, $B(L)$ have no common factors

    this is largely housekeeping, carrying around redundant factors that could be cancelled is a pain.
  - $A(L)$ and $B(L)$ have no unit roots.

    We want know where all the unit roots are so we can analyze them. Now we know there is just 1 unit root. Note that this assumption implies that there $C(L)$ has no unit roots
  - $A(L)$ has all roots outside unit circle

    Rule out explosive AR roots

▶ **Some observations**

- Difference x_t:

$$(1-L)x_t = (1-L)C(L)\varepsilon_t$$

- When differenced, x has a unit MA root; y does not.
- If we start from a series that is at most I(1), the theory says we can difference the series and look for a unit MA root.

  Presence of a unit MA root means the original was I(0)

- Language: When we difference an I(0) variable and thereby induce unit MA root, we say that the variable is overdifferenced (differenced more than needed to get to I(0))

▶ **More observations**

- The long run effect of a unit shock, ε_t, to x_t is zero

  That is $\lim_{j\to\infty} c_j = 0$

- The long run effect of a unit shock, ε_t, to $(1-L)y_t$ is zero
- Thus, the long-run effect of the shock on y_t is $C(1)$.

  The long-run effect on the level, y_t, is the sum of all the changes.

- (Note: we know $C(1) \neq 0$ because we assumed $A(L)$ and $B(L)$ have no unit roots.)
- That is, shocks have long-run effects on $I(1)$ variables but not on $I(0)$ variables.

▶ **Aside:: Deterministic trends**

- I could repeat the kind of analysis we've just done using process that are stationary around a deterministic time trend

  called trend stationary

- For example:

$$y_t = \alpha + \gamma t + \nu_t$$

  where ν_t stationary.

- First difference is

$$(1-L)y_t = \gamma + (1-L)\nu_t$$

  Thus, we get a series with a constant and a unit root in the MA part.

- If $\gamma = 0$, then the first difference is mean zero with a unit MA root.
- And so forth.
- Personal bias: I will not discuss trend stationarity.

  this is essentially never a sensible model in macro

- More precisely, use of deterministic trends has often led to silly results and never in my experience led to useful results.

- **Unit roots, do we know? Can we know?**
  - One way of describing an I(1) series is that shocks never die out entirely.
  - Thus, we can return to the Christiano-Eichenbaum question: Unit roots: Do We Know?

    (Actually, I think this could be rephrased, Can We know?)
  - You have 50 years of data. You want to know if shocks die out on an infinite horizon.
  - Put another way, you want to know if the correlation of y_t and y_s is zero when $|t-s|$ is arbitrarily large—say, spanning 10 centuries.
  - You won't and can't know, based on your data.
  - Without very strong restrictions on the way shocks die out, you can only reliably learn about stuff that happens multiple times in your sample.

    E.g. in your 50 year sample, you might learn about how much of shocks seems to remain after 5 years—you have about 10 nonoverlapping periods like this.
- **Thus: We don't know. We don't care. A nearly equivalent statement is that truth on the I(0), I(1) statement is overrated.**
- **Near observational equivalence**
  - Around 1990 a bunch of results were derived about 'near observational equivalence' of I(0) and I(1) processes

    This formalized, 'Can't know.'
  - A couple of my early papers were in this area

    e.g.,Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero; Faust, Jon; Econometrica, May 1999, v. 67, iss. 3, pp. 629-37. go

    http://www.jstor.org/stable/2999547
- **End of phase 1**
  - Phase 1, we learned that asymptotically unit roots are very different from stable roots
  - Then learned that in finite samples, the distinction is not so sharp.
  - Whether or not you have unit roots, however, unit root-like problems with statistics may appear whenever data are persistent.
  - The next phase (still going on) involved figuring out how best to proceed.
  - Before going on, however, it is worth reviewing a bit more unit root analytics.
- **Three bits of analytics**

- How do we prove the alternative asymptotic results sketched above?

  See the lecture on unit root asymptotics

- Implications of unit roots in the multivariate context: cointegration

  Lecture: cointegration

- Long-memory processes: $I(1)$, $I(0)$ and something in between.

  lecture: long-memory processes.