607
For completeness: long memory

Jon Faust

http://e105.org/e607

November 30, 2015

- **Both for completeness and to clean up an analytic detail we've glossed over, we now discuss long-memory processes**

- **Summability**
  - We have regularly written $C(L)$ and then referred to $C(1)$
  - Of course, $C(1)$ is the sum of the coefficients of the lag polynomial.
  - If $C(L) = \sum_j c_j L^j$ and our process is stationary, then by definition,

$$\sum c_j^2 < \infty$$

    (variance must be finite)

  - But this does not imply that $\sum |c_j| < \infty$

    that is, the sum, $C(1)$ may not converge to a finite number

- **Aside::** ℓ^2 **and** ℓ^1
  - We said before that the space of stationary MA representations is given by lag polynomials with coefficients in ℓ^2

    ℓ^2 is the square summable sequences

  - What we are adding now is a reminder that that ℓ^1 is strictly larger than ℓ^2.

- **Stationary, non-summable** $C(L)$**s**
  - So what is special about these stationary processes that don't have summable coefficients?
  - By definition, we know that the c_js are going to zero as j grows

- Under stationarity, they are going to zero fast enough that the sequence is square summable, but they may go to zero a bit too slow for the sequence to be absolutely summable.
- Thus, dependence dies out a bit more slowly for these processes than for absolutely summable processes
- This leads to the name: long-memory processes.
- Not sure I like this name. Square summable MAs can have memory as long as you like

  but, hey, I'm not head of the nomenclature division.

▶ **Long-memory processes**

- Absolutely (and, hence, square) summable processes have autocorrelation functions that, at sufficiently long lags, decay exponentially
- Square, but not absolutely, summable processes have autocorrelation functions that (at sufficiently log lags) decay hyperbolically.

  hyperbolically is slower than exponentially

▶ **Picture of autocorrelation function**

- Sorry, coming soon...

▶ **The spectrum**

- We have not talked a lot about the spectrum, but for completeness let me add this
- Absolutely summable processes have spectra that are finite at frequency zero and have a slope of zero at frequency zero.
- Square, but not absolutely summable processes, have spectra that are infinite at frequency zero and the slope near zero is obviously not zero.

▶ **Finally,**

- The square, but not absolutely, summable stationary processes can be written as having a fractionally differenced AR factor:

  $$(1-L)^d$$

  for $d \in (-1/2, 1/2)$
- Thus, we now have 3 choices: $I(0)$, $I(1)$ and in between: $I(d)$ $0 < |d| < 1/2$.
- Some folks proceeded by first deciding which of these three was relevant and then choosing the relevant asymptotic.

  Misguided for the same reason that the binary view was misguided.

▶ **Understanding fractional differencing**

- Suppose $(1-L)^d y_t = \varepsilon_t$.
- Write a Taylor series expansion of $\frac{1}{(1-L)^d} = (1-L)^{-d}$ to see the MA representation
- You'll find it has coefficients,

$$\psi_j = \frac{(j+d-1)!}{j!(d-1)!}$$

which leads to our claims about the autocorrelation function above.

▶ **Standard cites**

- J.R.M. Hosking, Fractional Differencing, Biometricag, v68 n1, Apr. 1981,165–176

  go

  http://www.jstor.org/stable/2335817

- C. W. J. Granger and Roselyne Joyeux, AN INTRODUCTION TO LONG-MEMORY TIME SERIES MODELS AND FRACTIONAL DIFFERENCING, Journal of Time Series Analysis,Volume 1, Issue 1, pages 15–29, January 1980

  go

  http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.1980.tb00297.x/

▶ **Do we know?**

- The intuition for why we don't care or know if something is short or long memory goes like this.
- The only difference between short and long memory is the rate of decay in the tail of the autocorrelation function.

  But in a finite sample, we don't learn about the tail (unless we make very strong assumptions).

▶ **Useless?**

- So is this elaboration useless?
- No.
- In practical terms, suppose we want to parsimoniously approximate the dynamics of a process with a slowly decaying autocorrelation function.
- We'd need a high order AR to match this pretty well for the first, say, 10 autocorrelations.
- But with one parameter (d), we could match it with a fractionally integrated process.
- Thus, we don't know or care about long vs. short memory.
- But we are pragmatic about approximating dynamics with a small number of parameters and fractionally differences processes can be useful in this regard.