Statistics 434: Bullet Points for Day 6
WRDS, CRSP, Real Asset Returns, and Normality Assumptions

The main task today is rock-bottom practical: How does one access the CRSP data via WRDS?

We also ask "When is it feasible to treat a series of returns as if they are normally distributed?" This question has both "straightforward" and "tricky" aspects. We do the "straightforward" part first and look at tools for deciding when a sequence of *independent identically distributed* observations may be reasonably considered to be normal. It is amusing to see what the classic normality tests when they are applied to real data.

We also take a theoretical look at the trickier question of trying to determine when normality is feasible for a sequence of returns that we only assume to be *stationary*. By a cruel example, we find that this question is literally impossible to answer — unless making a further assumption about our series. This leads us to the important notion of "ergodicity."

- Discussion of the in-coming HW2

- Background on Skewness and Kurtosis — Some Calibrating Theory

- Normality Tests for (Shapiro-Wilks Test and Jarque-Bera Test)

- Access to real returns

  1. WRDS Access
  2. Extracting CRSP holding period returns
  3. Creating an S-Plus data set and cleaning it up

- Exploration of real returns — especially the normality tests.

- Discussion of the out-going Homework 3 assignment

- Erogodicity — Stationarity's Little Brother

- Comments on the *Modeler's Paradox*

ASSUMPTIONS VS HYPOTHESIS — SCIENCE, ART, OR FLIM-FLAM

In elementary statistics, you are often told to assume that your observations are normally distributed. With this assumption (and others like identical distribution and independence) you get some powerful tools — such as the t-tests for sample means or regression coefficients.

Sometimes this is scientific; for example, mechanical observation errors are often decently normal (though usually with bias). Sometimes, this is wishful thinking which may or may not be useful, depending on the experience of the modeler and the intended use of the model. Sometimes it is an expository or "pedagogical" device — that's my favorite excuse. More often than one might hope — it's a hoax. The world has many analysts who do not seriously question their normal-based analyses. They just quietly hope that they won't get caught.