Maximum Likelihood Estimation

- A Fundamental principle of Estimation (as are OLS, GLS, IV, GMM estimation)

- Can be applied to a wide variety of Estimation problems (not just linear models)

- has excellent asymptotic properties when certain conditions hold

- Biggest drawback: Requires stronger distributional assumptions than other techniques.

Models estimated by ML must be fully specified.

Suppose we have a sample

$$\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_T)$$

The distribution of the sample is its jpdf. and in a fully specified model it has a known mathematical form

$$f(\underset{\sim}{y} \mid \underset{\sim}{\theta}) \quad \text{for } \theta \in \Omega$$

$\underset{\sim}{\theta}$ consists of k unknown parameters, Ω is the set of all possible values of θ (~~sample~~ parameter space). The specification of $f(\underset{\sim}{y} \mid \theta)$ and Ω represent a statistical model.

The basic goal of stat analysis is to collect samples, $\underset{\sim}{Y}$, that come from $f(\,)$ and use them to learn about $\underset{\sim}{\theta}$.

ML estimation

The basic principle is this: Given a well-defined, fully specified statistical model, what values of θ maximize the probability that my sample is being generated by the chosen $f(y|\theta)$?

So, we know $f(y|\theta)$, but θ is unknown.
and have a sample that is generated by it.

The likelihood function is algebraically identical to the j.p.d.f. of $\{Y_1, Y_2, \ldots Y_T\}$. It is denoted $L(\theta|y)$ where θ are the unknowns and y the realized values of sample

Maximize $L(\,)$ w/respect to θ. given y.

<u>Steps</u> 1. Write out j.p.d.f. of your model.

2. Call it the likelihood function

3. Maximize it with respect to θ.

This yields $\tilde{\theta}(\underset{\sim}{y})$, the MLE of $\underset{\sim}{\theta}$.
Since it depends on sample (which is random) it is a R.V. and has a p.d.f of its own.

Random Sampling of y's.

For many cases, the obs in the sample are thought to be statistically independent of one another. This makes the j.p.d.f. much easier to specify since it will be = to product of n marginal distributions

Let $f(y_t, \theta)$ denote the p.d.f of a typical obs, y_t. Then the j.p.d.f of the entire sample is

$$f(\underset{\sim}{y} \mid \underset{\sim}{\theta}) = \prod_{t=1}^{n} f(y_t \mid \theta)$$

As a product, this # can be ~~very large or~~ small. So, it is customary to work with its natural log

$\cancel{l(y \mid \theta) = ln f}$

$$\ell(\underset{\sim}{\theta}|\underset{\sim}{y}) = \ln\left[f(\underset{\sim}{y}|\theta)\right] = \sum_{t=1}^{T} \ell_t(\theta|\underset{\sim}{y})$$

where $\ell_t(\ )$ is the contribution to the log likelihood function made by obs. t. (which is $= \ln\left[f(y_t|\theta)\right]$

The subscripts are added to allow for the density to change from obs. to. obs.

Note: The natural log is a monotonic transformation $\Rightarrow$ the value of $\hat{\theta}$ that maximizes L will also maximize ℓ.

<u>Example</u>: Exponential

$$f(y_t|\theta) = \theta e^{-\theta y_t} \qquad y_t > 0,\ \theta > 0$$

Useful for dep vars that are > 0
waiting times, duration, Battery life, etc.

Take a T R.S., Take logs and add em up.

$$\ln[f(y_t|\theta)] = \ln\theta - \theta y_t$$

$$l = \sum_{t=1}^{n} (\ln\theta - \theta y_t) = n\ln(\theta) - \theta \sum y_t$$

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} - \sum y_t$$

$$\frac{n}{\tilde{\theta}} = \sum y_t \Rightarrow \tilde{\theta} = \frac{n}{\sum y_t}$$

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{n}{\tilde{\theta}^2} < 0 \quad \therefore \ \tilde{\theta} \text{ is } \underline{\underline{\text{max}}}.$$

Eval s.o.c. at MLE and make sure neg (or neg def)

If time dependent

$$f(\underset{\sim}{y}|\theta) = f(y_t | y_{t-1}, \ldots$$

y_t $\quad f(y_1, y_2) = f(y_2|y_1)\, f(y_1)$

$$f(y_3, y_2, y_1) = f(y_3 | y_2, y_1)\, f(y_2|y_1)\, f(y_1)$$

AND so on

you can build joint likelihood (P.d.f.) by taking product of conditionals.