Bayesian econometrics – introduction

Alternative way of thinking about estimation.
Allows the econometrician to take into account other information than what is in the data.
Bring extra information through a *prior* assumption about the parameters.
Difference Bayesian analysis and *classical*, or *frequentist* analysis partly philosophical

Classical estimation.

assume existence of *true* parameters θ.
Inference: Use data to get a *best estimate* (according to optimization criterion)
Example: estimate the mean of set of normal *iid* $\mathcal{N}(\mu, \sigma^2)$ observations

- μ
- σ

viewed as fixed, unknown numbers.
The estimator $\widehat{\theta}$ is a random variable
Find the best estimator based on

- Consistency/Unbiasedness. $\widehat{\theta} \to \theta$
- Mean squared error $E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)]$

Maximum likelihood estimation of normally distributed variables.
Suppose a variable x_i has a normal distribution with mean μ and variance σ^2.

1. Determine the maximum likelihood estimator of μ.
2. Determine the maximum likelihood estimator of σ^2.

Maximum likelihood estimation of normally distributed variables.
First recall the probability distribution for a normally distributed varible x_i

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

The likelihood function is

$$L(x;\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

We will instead of the likelihood function maximize the log-likelihood function:

$$\ell(y;\theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \sum_{i=1}^{n} -\frac{1}{2}\frac{1}{\sigma^2}(x_i-\mu)^2$$

Rewrite the log-likelihood function as

$$\ell(y;\theta) = \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \sum_{i=1}^{n} -\frac{1}{2}\frac{1}{\sigma^2}(x_i-\mu)^2$$

We find the estimator from the first order conditions, first estimating μ

$$\frac{\partial \ell}{\partial \mu} = \sum_i 2 \frac{1}{2\sigma^2}(y_i - \mu) = 0$$

$$\rightarrow \ \sum_i (y_i - \mu) = 0$$

$$\rightarrow \ \sum_i y_i = n\mu$$

$$\widehat{\mu}^{ml} = \frac{\sum_i y_i}{n}$$

and then estimating σ^2.

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} -\frac{1}{\sigma} - \sum_{i=1}^{n} \frac{1}{2}\left(\frac{0-(x_i-\mu)^2 2\sigma}{\sigma^4}\right)$$

$$0 = \sum_{i=1}^{n} -\frac{1}{\sigma} + \sum_{i=1}^{n} \frac{1}{2}\left(\frac{2(x_i-\mu)^2\sigma}{\sigma^4}\right)$$

$$0 = \sum_{i=1}^{n} -\frac{1}{\sigma} + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^3}\right)$$

$$0 = -\sum_{i=1}^{n} 1 + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$0 = -n + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$n = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\hat{\sigma}^2_{ml} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

Bayesian Estimation

Bayesian analysis treats everything in terms of probability distributions.
Even θ itself is viewed as a random variable.
Goal of Bayesian analysis:
Describe the analysts' uncertainty about θ in terms of a probability distribution.
Inference to a Bayesian analyst is to make probability statements.
Before observing the data, the analyst will summarize his knowledge in the *prior* distribution

$$f(\theta)$$

The *joint* probability density $f(y, \theta)$ describes the probability for observing both the data y and the parameter θ

$$f(y, \theta) = f(y|\theta) \cdot f(\theta)$$

The goal of Bayesian analysis is to find a *posterior density*

$$f(\theta|y),$$

The probability distribution of θ postulated after having observed the data.
Use the definition of conditional probability

$$f(\theta|y) = \frac{f(y,\theta)}{f(y)}$$

$$f(\theta|y) = \frac{f(y,\theta)}{\int_{-\infty}^{\infty} f(y,\theta)d\theta}$$

Using

$$f(y|\theta) = \frac{f(y,\theta)}{f(\theta)}$$

$$\rightarrow \quad f(y,\theta) = f(y|\theta)f(\theta)$$

Bayes formula:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int_{-\infty}^{\infty} f(y,\theta)d\theta}$$

Example: Bayesian estimation of mean of normal

$$y \sim \mathcal{N}(\mu, \sigma^2)$$

A Bayesian summarizes prior assumptions about the distribution into the parameters m and ν.

$$f(\mu|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2/\nu}} e^{-\frac{1}{2}\frac{(\mu-m)^2}{\sigma^2/\nu}}$$

The parameter m is the best guess on the mean based on prior knowledge, and the parameter ν describing the belief about *precision* of the prior knowledge.

The *sample* distribution

$$
\begin{aligned}
f(y|\mu,\sigma) &= \prod_{t=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{1}{\sigma^2}(y_t-\mu)^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{T}{2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}(\mathbf{y}-\mu\mathbf{1})(\mathbf{y}-\mu\mathbf{1})'}
\end{aligned}
$$

The goal is to produce the *posterior* distribution

$$f(\mu|y;\sigma)$$

The Bayesian estimator:

$$\widehat{\mu} = \left(\frac{\nu}{\nu + T}\right) m + \left(\frac{T}{\nu + T}\right) \frac{\sum_t y_t}{T}$$

– weighted average of

- prior m
- sample mean $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$.

The lower the precision ν, the more weight is put on the data, and the less on the prior.

Note that the limit when $\nu \to 0$ is the classical (sample average) estimate.
$\to$ *improper* or *diffuse* prior.

Bayesian analysis: discussion

Strength of Bayesian analysis: ability to account for *prior* information.

However: Reliance on prior leads to a *subjective* part – What is the source of the prior?

This subjectivity deterrent to the acceptance of Bayesian results.

However: classical analysis also using prior knowledge: The chosen probability model.

Bayesian statisticans – The Jehovas Witnesses of Statistics.

Bayesian analysis: discussion

Still routes for Bayesian results to get published.
For example, **?**, a way to ask if predictability matter.
Suppose an asset allocator is a Bayesian.
The predictability of monthly stock returns is small, by the usual (classical) inference (i.e. not statistically significant.)
However, one can argue that the predictability is important if it is *used* by the asset allocator.
Show that no matter what the prior, the historical data will still lead to the posterior distribution changing.
$\rightarrow$ the historical data *matters* for the asset allocator.

Shmuel Kandel and Robert F Stambaugh. On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, 51(2):385–424, June 1996.