Introduction to Econometrics

Brandon Lee

15.450 Recitation 9

Law of Large Numbers

- Suppose x_t are IID and $E[x_t] = \mu$. Then the Law of Large Numbers states that

$$plim \frac{1}{T} \sum_{t=1}^{T} x_t = \mu$$

  Intuitively, the LLN says that as the sample gets larger, the sample average approaches the true mean.

- The LLN is often the basis for establishing consistency of statistical estimators.

Consistency of OLS Estimator

- Suppose
$$y_i = x_i\beta + \varepsilon_i$$
where $E[\varepsilon_i|x_i] = 0$ (which then implies that the error term is uncorrelated to the sample: $E[x_i\varepsilon_i] = 0$)
- The OLS estimator is given by
$$\hat{\beta} = (X'X)^{-1} X'y$$
- Let's verify that $\hat{\beta}$ is consistent: that is, $plim\,\hat{\beta} = \beta$.

Continued

- Note

$$\begin{aligned}\hat{\beta} &= \left(X'X\right)^{-1} X'y \\ &= \left(X'X\right)^{-1} X' \left(X'\beta + \varepsilon\right) \\ &= \beta + \left(X'X\right)^{-1} X'\varepsilon\end{aligned}$$

Therefore,

$$\begin{aligned}plim\, \hat{\beta} &= \beta + plim\, \left(X'X\right)^{-1} X'\varepsilon \\ &= \beta + plim \left(\left(\frac{X'X}{N}\right)^{-1} \left(\frac{X'\varepsilon}{N}\right) \right)\end{aligned}$$

Continued

- The key here is that $\left(\frac{X'X}{N}\right)^{-1}$ will converge to some limit and so will $\left(\frac{X'\varepsilon}{N}\right)$. But by the Law of Large Numbers, we know that

$$plim\left(\frac{X'\varepsilon}{N}\right) = E\left[x_i\varepsilon_i\right] = 0$$

Therefore,

$$plim\,\hat{\beta} = \beta$$

Central Limit Theorem

- Suppose that x_t is a random vector such that $E[x_t] = \mu$ and $Var(x_t) = \Omega$. The Central Limit Theorem states that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu) \Rightarrow N(0, \Omega)$$

  Here, the convergence is in "convergence in distribution".
- The CLT is often used to derive asymptotic distribution of statistical estimators.

Maximum Likelihood Estimator

- Having observed the sample $x_1, \ldots, x_T$, we want to estimate the unknown true parameter θ_0 of the data generating process $f(x;\theta)$.
- Maximum likelihood estimation is an intuitive procedure in which the probability of observing our sample is maximized at our maximum likelihood estimate $\hat{\theta}_{MLE}$.
- Likelihood function (it is a function of the parameter θ, taking as given the sample): $L(\theta | x_1, \ldots, x_T)$
- Log-likelihood function (this is typically what we work with): $\mathscr{L}(\theta | x_1, \ldots, x_T)$
- The goal is to find $\hat{\theta}$ that maximizes our (log-)likelihood function. Sometimes we can do this by finding a solution to the first order condition, but in other situations we may have to resort to numerical optimization routines.

Example: Mixture of Normals

- Assume that asset returns are IID and normally distributed, $N\left(\mu, \sigma^2\right)$. We've seen in the lectures that the MLE of μ and σ^2 are simply given by the sample mean and sample variance, respectively.
- Let's assume instead that returns are IID over time, but now drawn from a mixture of normal distributions: that is with probability λ, it is drawn from $N\left(\mu_1, \sigma_1^2\right)$ and with probability $1-\lambda$, it is drawn from $N\left(\mu_2, \sigma_2^2\right)$. This is one of the popular approaches to modelling fat-tail distributions.
- Now the parameters of the model are $\left(\lambda, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right)$.

Continued

- Note that

$$f\left(R_t|\lambda,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2\right)$$
$$=\lambda\cdot\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(R_t-\mu_1)^2}{2\sigma_1^2}}+(1-\lambda)\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{(R_t-\mu_2)^2}{2\sigma_2^2}}$$

and since we have IID sample,

$$L\left(\lambda,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2|R_1,\ldots,R_T\right)=\prod_{t=1}^{T}f\left(R_t|\lambda,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2\right)$$

The log-likelihood function is given by

$$\mathscr{L}\left(\lambda,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2|R_1,\ldots,R_T\right)$$
$$=\sum_{t=1}^{T}\log\left(\lambda\cdot\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(R_t-\mu_1)^2}{2\sigma_1^2}}+(1-\lambda)\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{(R_t-\mu_2)^2}{2\sigma_2^2}}\right)$$

Example: GARCH

- Suppose $R_t \sim N\left(\mu, \sigma_t^2\right)$. The interesting aspect of this specification is time-varying volatility. In particular, we assume GARCH(1,1) structure:

$$\sigma_t^2 = \alpha + \beta\left(R_{t-1} - \mu\right)^2 + \gamma\sigma_{t-1}^2$$

  We have in mind $\beta > 0$ and $\gamma > 0$ so that past realized and latent volatility carry over to the current period. These kinds of specifications can capture the volatility clustering we see in the data.
- The parameters of our model are $\left(\mu, \alpha, \beta, \gamma, \sigma_0^2\right)$.

Continued

- The likelihood function is given by

$$L\left(\mu,\alpha,\beta,\gamma,\sigma_0^2|R_1,\ldots,R_T\right)=\prod_{t=1}^{T} f\left(R_t|\mu,\alpha,\beta,\gamma,\sigma_0^2;R_1,\ldots,R_{t-1}\right)$$
$$=\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_t^2}}e^{-\frac{(R_t-\mu)^2}{2\sigma_t^2}}$$

  Note that σ_t^2 is included in the information set $(R_1,\ldots,R_{t-1})$.
- Optimizing this objective function cannot be done analytically because evolution of σ_t^2 depends on all the parameters in a non-trivial manner. We have to resort to numerical methods to find the optimum.

15.450 Analytics of Finance

Fall 2010