Econometrics: Time series basics

Large sample theory

Compiled by the PGDM Research and Analytics cell

0.1 Large Sample theory

CLRM models are based on finite sample assumptions and they tend to break down if these assumptions are violated. The key difference of this with the Large sample theory approach is that **rather than making assumptions on the sample of a given size, large sample theory makes assumptions on the stochastic process that generates the samples**. Before delving into this topic we must review the essentialy limit theorems.

0.1.1 Limit theorems

In this section we look at the limiting behaviour of a sequence of random variables $(z_1, z_2, \cdots, z_n)$. Such a sequence is often denoted as $\{z_n\}$. We now look at some popular modes of convergence.

- A sequence of random scalars $\{z_n\}$ **converges in probability** to a constant α if for any $\epsilon > 0$ we have:

$$\lim_{n\to\infty} P(|z_n - \alpha| > \epsilon) = 0 \tag{1}$$

  This constant α is known as the probability limit of the sequence $\{z_n\}$ and is often denoted as $plim_{n\to\infty} z_n = \alpha$. We can also write this as:

$$z_n \xrightarrow{P} \alpha \tag{2}$$

  We can actually extend this concept to denote convergence for a family of random variable sequences denoted as **random vectors** which may be written as K dimensional vectors $\{\boldsymbol{z_n}\}$, implying element by element convergence to a vector of constants $\boldsymbol{\alpha}$. So we can define convergence for the k^{th} element of such a vector for any $\epsilon > 0$ as:

$$\lim_{n\to\infty} P(|z_{nk} - \alpha_k| > \epsilon) = 0 \tag{3}$$

- A sequence of random scalars $\{z_n\}$ tends to converge **almost surely** to a constant α if we have:

$$P\left(\lim_{n\to\infty} z_n = \alpha\right) = 1 \tag{4}$$

$$z_n \overset{a.s}{\to} \alpha \tag{5}$$

- A sequence of random scalars $\{z_n\}$ tends to **converge in mean square** to a constant α is we have:

$$\lim_{n\to\infty} E[(z_n - \alpha)^2] = 0 \tag{6}$$

  Whereas all the above convergence definitions involve convergence towards a constant, they can analogously be defined for random variables as well.

- Now let $\{z_n\}$ be a seqence of random scalars and F_n as the cumulative distribution of z_n. We say that the sequence $\{z_n\}$ **converges in distribution** to a random variable/scalar z if the CDF of z_n denoted as F_n converges to the CDF of F of z at every continuity point of F. Note that F is called the **asymptotic** or limiting distribution of z_n. This is denoted as:

$$z_n \overset{D}{\to} z \tag{7}$$

0.2 Fundamental concepts in time series

Note that a **stochastic process** is nothing but a sequence or a family of random variables, whose behaviour (especially limiting) we are often interested in. Now if the index of the random variables is interpreted as representing time, then the stochastic process is called a **time series**. Note that if $\{z_i\}$ is a time series, then its **realization** or **sample path** is the assignement for each i, a possible value that z_i might take. Therefore a realization of $\{z_i\}$ is just a sequence of real numbers.

0.2.1 Ergodic stationarity

Now we say that the fundamental problem of a time series is that we can observe its realizations only once. For example, the sample of historic inflation rate values between say 1946 to 1995 is nothing but a string of 50 numbers, which happens to be but one realized outcome of the underlying stochastic process for the inflation rate family of time indexed random variables. In an alternate historical outcome, we would have obtained a different sample. Now suppose that we could observe the same historical period many times, in such case we would be able to observe many samples of the string of 50 numbers. Therefore now if we want the mean inflation rate for the year 1995 we could simply compute the average of the 50^{th} number across each sample. Such a population mean estimation is called the **ensemble mean**. Basically the ensemble mean is the **average across all possible states of nature for a given instance of time**. Now it is quite obvious that we cannot observe multiple histories, however we can formulate our methodology a bit differently. Now suppose we assume that the distribution of inflation rate remains unchanges across time. In this case the 50 numbers we observe can be thought of as derived from effectively the same distribution. Therefore in such a situation the time average of a single string would be a consistent estimator for the ensemble mean.

0.2.2 Stationary process

A stochastic process $\{z_i\}$ is said to be **strictly stationary** is for any given integer r and a set of subscripts $i_1, i_2, \cdots, i_r$ the joint distribution of $(z_i, z_{i_1}, z_{i_2}, \cdots, z_{i_r})$ depends only on $(i - i_1), (i_2 - i_1), \cdots, (i_r - i)$ but not on i itself. We can therefore say that the distribution of (z_1, z_5) is effectively the same as (z_{12}, z_{16}). We can say that **relative position in the sequence** is what matters for the distribution. Also since the distribution of z_i does not depend on i, we say that the mean, variance and all other higher moments remain the same across values of i. Further we note that random variables that typically exhibit some form of a time trend, are said to be **nonstationary**. Note that the process of reducing time series with trend to a stationary series is called **trend stationary**. A process is said to be trend stationary if it becomes stationary after subtracting from it a function of time (possibly in the form of a time indexed variable). If a process is not stationary but its first difference given by:

$$z_i - z_{i-1} \tag{8}$$

is stationary, then we call $\{z_i\}$ as **difference stationary**.

0.2.3 Covariance stationary process

A process is said to be **weakly stationary** if we have:

- $E(z_i)$ does not depend on i
- $Cov(z_i, z_{i-j})$ exists and depends only on j but not on i. Therefore we can say that $Cov(z_1, z_5) = Cov(z_{12}, z_{16})$.

Now considering our random variables as random vectors instead (denoting joint random variables of K dimension), we can compute the j^{th} order **autocovariance** denoted by $\boldsymbol{\Gamma_j}$ as follows:

$$\boldsymbol{\Gamma_j} = Cov(\boldsymbol{z_i}, \boldsymbol{z_{i-1}}) \tag{9}$$

Note that the term 'auto' signifies that we are taking the random variables from the same process. Further, covariance stationarity implies that this autocovariance does not depend on i and satisfies:

$$\boldsymbol{\Gamma_j} = \boldsymbol{\Gamma'_{-j}} \tag{10}$$

Note that the 0 order autocovariance is nothing but the variance, which is constant across time.

$$\boldsymbol{\Gamma_0} = Var(\boldsymbol{z_i}) \tag{11}$$

Now for a scalar covariance stationary process $\{z_n\}$, the j^{th} order autocovariance is given by:

$$\gamma_j = \gamma_{-j} \tag{12}$$

If we take a string of n successive values $(z_i, z_{i+1}, \cdots, z_{i+n-1})$ from a scalar process, then by covariance stationarity, its $n \times n$ variance-covariance matrix is the same as that of $(z_1, z_2, \cdots, z_n)$ and is given below. Note that this shows the **autocovariance matrix**.

$$Var(z_i, z_{i+}, \cdots, z_{i+n-1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \tag{13}$$

With this defined, we can write the j^{th} order **autocorrelation coefficient**, ρ_j as follows:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{Cov(z_i, z_{i-j})}{Var(z_i)} \tag{14}$$

If we plot the sequences $\{\rho_j\}$ along with the points $j = 1, 2, \cdots$ we would get a **correlogram**.

0.2.4 White noise process

An important class of weakly stationary stochastic processes that have zero mean and exhibit no serial correlation are called **white noise series**.

$$E(\boldsymbol{z_i}) = 0 \tag{15}$$

$$Cov(\boldsymbol{z_i}, \boldsymbol{z_{i-j}}) = 0 \tag{16}$$

Note that an iid sequence with zero mean and finite mean is a special case of a white noise series.

0.2.5 Ergodicity

A stationary process $\{z_n\}$ is said to be **ergodic** is for any two bounded functions $f : R^k \to R$ and $g : R^l \to R$ we have:

$$\lim_{n\to\infty} |E[f(z_i, \cdots, z_{i+k})g(z_{i+n}, \cdots, z_{i+n+l})]| \tag{17}$$

$$= |E[f(z_i, \cdots, z_{i+k})]||E[g(z_{i+k}, \cdots, z_{i+n+l})]| \tag{18}$$

We say that a stationary process is ergodic if it is **asymptotically independent**. This means that two random variables positioned far apart in the sequence are almost independently distributed. A stationary process that is ergodic is called **ergodic stationary**.

References

[1] Fumio Hayashi - Econometrics