B-74-3-B Time Series Econometrics

Fabrizio Iacone
Università degli Studi di Milano

Academic Year 2019-2020

Chapter 1 - Introduction, Stationarity and Ergodicity

TIME SERIES

Statistical analysis of data observed over time

The data:

- observed between two dates, normalised as $t = 1$ and $t = T$;
- Equispaced,
  i.e. we observe $Y_1, Y_2, ..., Y_t, Y_{t+1}, ..., Y_{T-1}, Y_T$
  and no intermediate observation is missing
- Y_t depends on Y_s if $s < t$
- Y_t does not depend on Y_s if $s > t$

then, the vector

$$\{Y_1, Y_2, ..., Y_t, Y_{t+1}, ..., Y_{T-1}, Y_T\}'$$

is a time series.

Moments

For a generic random variable we can define MEAN, VARIANCE, and for pairs of random variables we can define COVARIANCE, CORRELATION...

In a time series we define these for each Y_t:

- Mean: $E(Y_t) = \mu_t$
- Variance: $E\left[(Y_t - \mu_t)^2\right] = \sigma_t^2$
- Covariance: $E\left[(Y_t - \mu_t)(Y_{t+j} - \mu_{t+j})\right] = \gamma_t(j)$
- Correlation: $\frac{\gamma_t(j)}{\sigma_t \sigma_{t+j}} = \rho_t(j)$

Operators

Lag operator: L

$$LY_t = Y_{t-1}$$

So,

$$\begin{aligned} L^2 Y_t &= L\left(LY_t\right) = L\left(Y_{t-1}\right) = Y_{t-2} \\ L^{-1} Y_t &= Y_{t+1} \end{aligned}$$

First Difference operator: $\Delta = 1 - L$

$$\begin{aligned} \Delta Y_t &= (1 - L)\, Y_t = Y_t - LY_t \\ &= Y_t - Y_{t-1} \end{aligned}$$

Also,

$$\begin{aligned} \Delta^2 Y_t &= (1 - L)^2\, Y_t \\ &= \left(1 - 2L + L^2\right) Y_t \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \end{aligned}$$

Stationarity and ergodicity

$\{Y_1, ..., Y_T\}'$ is a single realisation from a stochastic process $\{Y_t\}_{t=-\infty}^{\infty}$.

We are interested in the model that generated the time series, but we do not know it. How can we make inference, using one single observation?

We must use the fact that this is a $T-$dimensional observation:

- Restrict heterogeneity over time
- Restrict dependence over time

Restrict heterogeneity
Assume that some properties are common to all the Y_t in $\{Y_1, ..., Y_T\}'$.
For example,

Covariance Stationarity

$$E(Y_t) = \mu \ \forall t$$

$$E[(Y_t - \mu)(Y_{t+j} - \mu)] = \gamma(j) \ \forall t$$

i.e. the first two moments do not depend on the position of Y_t.

In this way, we may try to estimate μ or $\gamma(j)$ using the sample counterparts.

"Covariance stationarity" is also known as "Weak stationarity" or simply as "Stationarity" (without other references).

For stationary processes, we shorten the notation and introduce

$$\gamma_j \text{ for } \gamma(j), \ \rho_j \text{ for } \rho_t(j)$$

to indicate the autocovariance and autocorrelations, respectively.

The plot of γ_j (against j) is called autocovariance function.
The plot of ρ_j (against j) is called autocorrelation function.

An alternative restriction on heterogeneity is **Strict stationarity**:

for any $j_1, ..., j_n$, the distribution of $\left\{Y_{t+j_1}, ..., Y_{t+j_n}\right\}'$ and of $\left\{Y_{t+\tau+j_1}, ..., Y_{t+\tau+j_n}\right\}'$ is the same for any τ.

Strict and Covariance stationarity do not imply each other.

Sufficient condition for stationarity

White Noise Process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$, $E(\varepsilon_t \varepsilon_\tau) = 0$ if $t \neq \tau$, is called white noise.

Consider process $\{Y_t\}_{t=-\infty}^{\infty}$ defined as

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

If

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty$$

and ε_t is white noise, then Y_t is stationary.

Restrict dependence over time

It may be that for n large enough Y_t and Y_{t+n} may be treated as independent, in which case the standard statistical theory would apply. If the process is stationary, then, the sample moments would estimate the population moments consistently.

We may generalise this argument and allow for some dependence, provided that it is not too much, and it vanishes quickly as n gets large.
One theoretical restriction that makes dependence vanish quickly as n gets large is called MIXING. At our stage, we may think *mixing* as asymptotic independence, even though the definition is more complex.

ERGODICITY (heuristic)

The property that we can *successfully* estimate the properties of a process from a long time series is called ergodicity.

Suppose that we have N identical processes, all observed at the same point in time, one time series of dimension N. A process is ergodic when the average of the observations at time t of the N identical processes is the same (in a probabilistic sense) a the average of the N dimensional time series.

For example, a stationary process is ergodic for the mean in mean squared sense (MS) if the sample average estimates (MS) consistently the expected value.

We can understand this with two counterexamples (from Wikipedia)

- *Suppose that we have two coins: one coin is fair and the other has two heads. We choose (at random) one of the coins first, and then perform a sequence of independent tosses of our selected coin. Let X[n] denote the outcome of the nth toss, with 1 for heads and 0 for tails. Then the ensemble average is $1/2(1/2+1)=3/4$; yet the long-term average is 1/2 for the fair coin and 1 for the two-headed coin. So the long term time-average is either 1/2 or 1. Hence, this random process is not ergodic in mean.*
- *An unbiased random walk is non-ergodic. Its expectation value is zero at all times, whereas its time average is a random variable with divergent variance.*

Some examples

1. (Model MD) Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be an independent, normally distributed process, with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$, and consider process $\{Y_t\}_{t=-\infty}^{\infty}$ defined as

$$Y_t = \varepsilon_t \varepsilon_{t-1}$$

Then, $\{Y_t\}_{t=-\infty}^{\infty}$ is white noise (and therefore stationary) but not independent. Moreover, Y_t is independent from Y_{t+n} for $n \geq 2$, and $\{Y_t\}_{t=-\infty}^{\infty}$ is therefore ergodic.

2. (Model MA1) Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be an independent, identically distributed process, with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$, and consider process $\{Y_t\}_{t=-\infty}^{\infty}$ defined as

$$Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

Then, $\{Y_t\}_{t=-\infty}^{\infty}$ is stationary. Moreover, Y_t is independent from Y_{t+n} for $n \geq 2$, and $\{Y_t\}_{t=-\infty}^{\infty}$ is therefore ergodic.

3. (Model AR1) Let $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ be an independent, normally distributed process, with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$, and consider process $\{Y_t\}_{t=-\infty}^{\infty}$ defined as

$$Y_t = c + \phi Y_t + \varepsilon_t$$

and $|\phi| < 1$.
Then, using repeated substitution,

$$Y_t = \frac{c}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

and $\{Y_t\}_{t=-\infty}^{\infty}$ is therefore stationary.

To check mixing, recall that, for two normally distributed random variables Y, X, with $E(Y) = \mu_Y$, $Var(Y) = \sigma_Y^2$, $E(X) = \mu_X$, $Var(X) = \sigma_X^2$, $Cor(Y, X) = \rho$, the joint density $f_{Y,X}(y, x)$ is

$$
\begin{aligned}
f_{Y,X}(y,x) &= \frac{1}{2\pi\sigma_Y\sigma_X\sqrt{1-\rho^2}} \\
&\qquad exp\{-\frac{(\frac{y-\mu_Y}{\sigma_Y})^2 - 2\rho(\frac{y-\mu_Y}{\sigma_Y})(\frac{x-\mu_X}{\sigma_X}) + (\frac{x-\mu_X}{\sigma_X})^2}{2(1-\rho^2)}\}
\end{aligned}
$$

Then, noticing $Cor(Y_t, Y_{t+n}) = \phi^n$, $Cor(Y_t, Y_{t+n}) \rightarrow 0$ the joint density for Y_t, Y_{t+n} is such that

$$f_{Y_t, Y_{t+n}}(y_t, y_{t+n}) \rightarrow f_{Y_t}(y_t) f_{Y_{t+n}}(y_{t+n})$$

(this is a heuristic check of mixing and ergodicity)