UiO : University of Oslo

ECON 4160: Econometrics-Modelling and Systems Estimation

Lecture 5: Stationary time series

Ragnar Nymoen

Department of Economics University of Oslo

10 September 2018

The reference to this lecture is:

- Chapter 4 in the textbook: *Dynamic Econometrics for Empirical Macroeconomic Modelling*.

Stochastic process and time series

- A stochastic process is a collection of random variables
- A time series is a stochastic process where the random variables are ordered by the time index t.
- In this course we study time series that are generated by stochastic difference equations.
- This means that we can use the solution of a difference equations to obtain expressions for the moments of the time series variables: expectation, variance and (auto)covariance.
- ...and to obtain an inference theory for time series.
- Main point today: Equivalence between stable solution and stationarity.

Examples of time series

- A time series which is defined by a stochastic difference equation is:

  $$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t,$$

  where:

  $$\varepsilon_t \sim IID\left(0, \sigma_\varepsilon^2\right) \;\; \forall\, t, \tag{1}$$

  exactly as in the previous lectures.
- Two other examples:

  $$Y_t = t + \varepsilon_t, \;\; \varepsilon_t \sim IID\left(0, \sigma_\varepsilon^2\right), \text{ and} \tag{2}$$

  $$\Delta Y_t = 1 + \Delta \varepsilon_t, \tag{3}$$

Autocovariance

- Stationarity is a characteristic of the linear properties of the process, namely expectation, variance and (auto)covariance.
- For the time series Y_t, we define the *autocovariances* $\gamma_{j,t}$ as:

$$\gamma_{j,t} = E[(Y_t - \mu_t)(Y_{t-j} - \mu_t)], \ j = 0, 1, 2, \dots, \qquad (4)$$

where the expectation μ_t is defined in the standard way:

$$\mu_t = E(Y_t) = \int_{-\infty}^{\infty} Y_t f_{Y_t}(Y_t) dY_t,$$

for a density function $f_{Y_t}(Y_t)$.

Stationarity (definition)

Definition (4.1 (in book) Weak stationarity)

if $E(|Y_t|^2) < \infty \; \forall \; t$, and if neither μ nor γ_j, depend on t, the Y_t-process $\{Y_t; t = 0, \pm 1, \pm 2, \pm 3, ...\}$ is covariance stationary, or weakly stationary:

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

and

$$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \forall \; t, \; j.$$

- $j = 0$: γ_0 is the variance of Y_t: $Var(Y_t) = \gamma_0$, which does not depend on time if Y_t is a stationary time series.
- $j \neq 0$: $\gamma_j = \gamma_{-j}$ if Y_t is stationary. Meaning?

Autocorrelation function (ACF)

The theoretical *theoretical autocorrelation function*, ACF, $\{\zeta_{t1}, \zeta_{t2}, \ldots\}$ is defined as:

$$\zeta_{j,t} = \frac{Cov(Y_t, Y_{t-j})}{Var(Y_t)} = \frac{\gamma_{j,t}}{\gamma_{0,t}}. \tag{5}$$

For an actual (observed) time series realization $\{Y_t; t = 1, 2, 3, ...T\}$, the *empirical ACF* becomes:

$$\hat{\zeta}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}. \tag{6}$$

where $\hat{\gamma}_j$ denotes empirical second order moments:

$$\hat{\gamma}_j = 1/T \sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}),\ j = 0, 1, 2, \ldots, T-1 \tag{7}$$

Estimation of ACF

- For stationary processes, $\hat{\zeta}_j$ $(j = 0, 1, 2, \ldots)$ are consistent estimators of the theoretical autocorrelations.
- A test of whether peaks in $\hat{\zeta}_j$ are significantly different from zero, can be based on the interval $0 \pm z_{\alpha/2}/\sqrt{T}$ where $z_{\alpha/2}$ is the (critical) value of the standard normal variable z with $P(|z|) > z_{\alpha/2} = \alpha$.

White noise

A process $\{\varepsilon_t; t = 0, \pm 1, \pm 2, \pm 3, ...\}$ is called *white noise* if:

$$E(\varepsilon_t) = 0 \tag{8}$$

$$Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2 \tag{9}$$

$$Cov(\varepsilon_t, \varepsilon_{t-j}) = \gamma_j = 0 \text{ for } j \neq 0. \tag{10}$$

- It is common to denote $\varepsilon_t \sim IID\left(0, \sigma_\varepsilon^2\right)$ in (1) as white noise.
- If the identical distribution is normal, we say that ε_t is *Gaussian white noise*, or (shorter), a *Gaussian process*:

$$\varepsilon_t \sim IIN(0, \sigma_\varepsilon^2) \text{ for all } t.$$

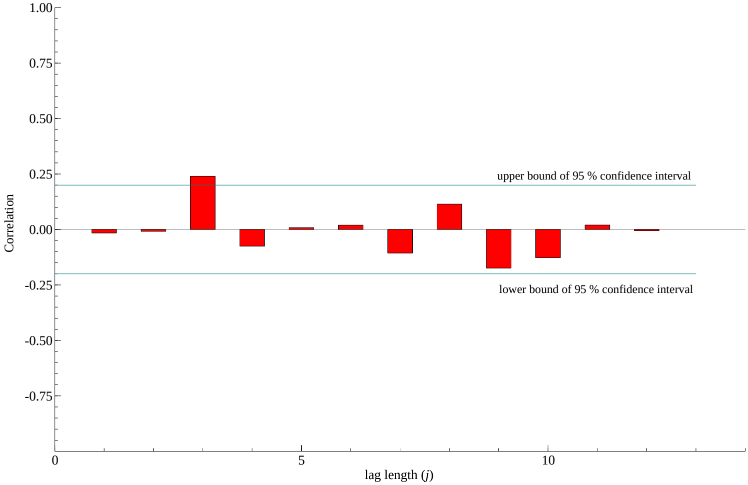

Figure 1: Empirical *ACF*, of a white noise series. The estimation is based on computer generated data with $\sigma_\varepsilon^2 = 1$ and $t = 1, 2, ..., 100$.

White noise signal and sound

- White noise on Wikipedia
- White noise generator
- White noise is the sound counterpart to *white light*

Linear filter

- A linear filter is a linear combination of L^k, $k = 0, \pm 1, \pm 2, \pm 3, \ldots$
- If two linear filters are multiplied, the result is a new linear filter.
- The two linear filters $a(L)$ and $b(L)$, and a time series X_t, can give a new time series Y_t as:

$$Y_t = a(L)b(L)X_t = c(L)X_t,$$

where the multiplication:

$$c(L) = a(L)b(L),$$

is carried through as if the lag operator L is a number.

Preservation of stationarity through linear filtering

- Linear filters can be infinite (also called a symmetric filter), or finite as here:

$$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j. \tag{11}$$

- A *well behaved linear filter* is characterized by:

$$\sum_{j=-\infty}^{\infty} \left|\psi_j\right| < \infty$$

- (another convention is: $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$.)
- Theorem 4.1: If $\psi(L)$ is well behaved, a filtering of a stationary process by $\psi(L)$ will give a new stationary process.

AR(1) process

- The time series $\{Y_t; t = 0, \pm 1, \pm 2, \pm 3, ...\}$ is an $AR(1)$-process when it is defined by the difference equation:
$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t, \tag{12}$$
where ϕ_0 and ϕ_1 are constant parameters, and ε_t is white noise with variance σ_ε^2.
- Solution, assuming that Y_0 is known:
$$Y_t = \phi_0 \sum_{j=0}^{t-1} \phi_1^i + \phi_1^t Y_0 + \sum_{j=0}^{t-1} \phi_1^i \varepsilon_{t-j}. \tag{13}$$
- and, solution assuming $-1 < \phi_1 < 1$:
$$Y_t = \frac{\phi_0}{1 - \phi_1} + \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}. \tag{14}$$
- Use (14) to show that Y_t is stationary if $-1 < \phi_1 < 1$.

- In Ch 4.5.3 in the book we take a different route first, and shows that Y_t is stationary if $-1 < \phi_1 < 1$, by starting from (13) and deriving:
  - $E(Y_t \mid Y_0)$ and $E(Y_t)$
  - $Var(Y_t \mid Y_0)$ and $Var(Y_t)$
  - ACF: $\zeta_j = \phi_1 \zeta_{j-1}$ for AR(1).
- ...the point being that expectation, variance and autocorrelations do *not* depend on time if $-1 < \phi_1 < 1$.
- Note that the condition for stationarity of the series Y_t is the same as for the conditions for (asymptotic global) stability of the solution of the difference equation for Y_t.
- This equivalence also holds for more general time series.

ARMA(1,1)

- Autoregressive Moving Average (ARMA) are combined processes.
- The simplest model is ARMA(1,1):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, t = 1, 2, \ldots, T, \qquad (15)$$

  where ϕ_0, ϕ_1 and θ_1 are constant parameters and ε_t is white noise with variance σ_ε^2.
- $Y_t \sim ARMA(1,1)$ is stationary if $-1 < \phi_1 < 1$
- Why ? Because the solution for Y_t then can be written as a well behaved filtering of $\varepsilon_t, \varepsilon_{t-1}, ...$
- The value taken by θ_1 does not come into consideration: The MA part belongs to the non-homogenous part of the difference equation, while stability hinges on the coefficients (characteristic roots) of the homogenous part (ie the AR-part).

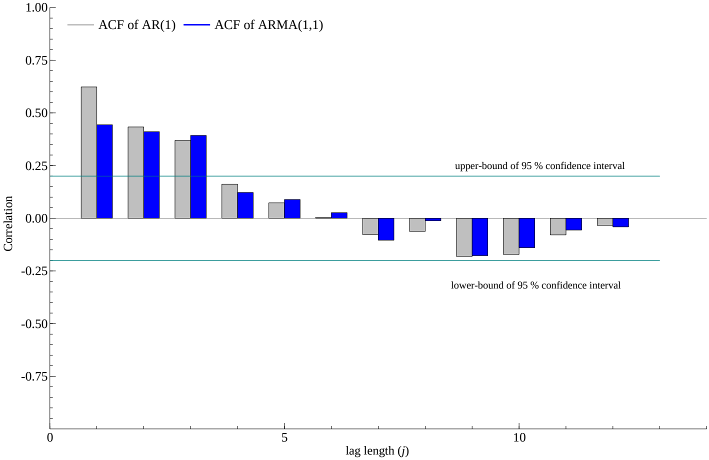

Figure 2: Empirical *ACF* of an AR(1) process ($\phi_1 = 0.6$), and of an ARMA(1,1), ($\phi_1 = 0.6$, $\theta_1 = -0.3$). The estimation is based on computer generated data, $t = 1, 2, ..., 100$, with $\sigma_\varepsilon^2 = 1$, for both processes.

Generalization to ARMA(p,q)

Definition (ARMA)

Let $\{Y_t; t = 0, \pm 1, \pm 2, \pm 3, \ldots .\}$ denote a time series, and let $\{\varepsilon_t\}$ denote a white noise process. $Y_t \sim ARMA(p,q)$ is given by:

$$\pi(L)Y_t = \phi_0 + \theta(L)\varepsilon_t \tag{16}$$

where $\pi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$.

$\pi(L)$ was defined in (3.76) in Ch. 3.5.2. It can be expressed compactly as

$$\pi(L) = 1 - \phi(L)$$

with the definition of $\phi(L)$ given in (3.77).

Theorem 4.2, 4.3 and 4.4 say that:

- $Y_t \sim ARMA(p,q)$ is covariance stationary if and only if the all the roots of:
$$\lambda^p - \phi_1\lambda^{p-1} - \ldots - \phi_p = 0$$
are different from one in magnitude.
- If and only if all the characteristic roots are less than one in magnitude, the stationary solution of $Y_t \sim ARMA(p,q)$ is a one-sided filter of the form:
$$Y_t - \mu = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{17}$$
where:
$$\mu = \frac{\phi_0}{\pi(1)} = \frac{\phi_0}{1-\phi(1)},$$
- For this model, referred to as *causal ARMA*, the stationary solution is found by backward substitution.

ML Estimation of AR(1)

- In Lecture 2 we reviewed the log-likelihood function for the regression model equation under the "classical" IID assumptions.
- For $Y_t \sim AR(1)$ independence is an irrelevant assumption (except for $\phi_1 = 0$)
- Instead we must regard for example Y_3, Y_2, Y_1 as *dependent variables* with joint pdf:

$$f_{Y_3,Y_2,Y_1}(y_3, y_2,y_1)$$

- which can be factorized by sequential conditioning:

$$\begin{aligned} f_{Y_3,Y_2,Y_1}(y_3, y_2,y_1) &= f_{Y_3|Y_2}(y_3 \mid y_2)\cdot f_{Y_2,Y_1}(y_2,y_1) \\ &= f_{Y_3|Y_2}(y_3 \mid y_2)\cdot f_{Y_2|Y_1}(y_2 \mid y_1)\cdot f_{Y_1|Y_0}(y_1 \mid y_0) \end{aligned}$$

- can do this for $Y_T, Y_{T-1}, ..., Y_1$

- If we assume normality, the conditional pdf's are normal, and the situation is exactly the same as with "ordinary regression"
- Can therefore formulate the log-likelihood function in the same way, to become:

$$\mathcal{L}(\phi_0, \phi_1, \sigma^2 \mid Y_0) = -\frac{T}{2}(\ln(2\pi/\sigma^2)) - \sum_{t=1}^{T} \frac{(Y_t - \phi_0 - \phi_1 Y_{t-1})^2}{2\sigma^2}, \tag{18}$$

- The only change from the IID case is that, strictly speaking, since we condition on a realization of Y_0, we may speak of conditional log likelihood.
- In practice, this conditioning has little importance for the results.

ML estimators

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} Y_t (Y_{t-1} - \bar{Y})}{\sum_{t=1}^{T} (Y_{t-1} - \bar{Y})^2}, \tag{19}$$

and

$$\hat{\phi}_0 = \bar{Y} - \hat{\phi}_1 \bar{Y}, \tag{20}$$

where it is understood that $\bar{Y}$ is the mean of the lagged process $\{Y_{T-1}, Y_{T-2}, \ldots, Y_0\}$.
For the variance:

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 = T^{-1} \sum_{t=1}^{T} [\underbrace{(Y_t - \hat{\phi}_0 - \hat{\phi}_1 Y_{t-1})}_{\hat{\varepsilon}_t}]^2. \tag{21}$$

Properties of estimators and t-values

- Large samples properties (from asymptotic theory):
  - The estimators are consistent
  - t-values are approximately normal (so conventional procedure for hypothesis testing can be used).
- Small sample properties
  - The estimators must be biased, since Y_{t-1} is not strictly exogenous but predetermined.
  - For the case of $\phi_0 = 0$, it has been shown that:

$$E(\hat{\phi}_1 - \phi_1) \approx \frac{-2\phi_1}{T} \tag{22}$$

    which is known as the *Hurwicz bias*.
  - Monte Carlo simulation is suited for getting a "feel for the bias"

Generalizations

- The ML interpretation of OLS generalizes directly to AR(p)
  - The conditioning of the likelihood is then on $Y_0, Y_{-1}...$, but that is no principal change.
- Also extends to ARMA, but maximization then requires numerical iteration, Non-linear Least Squares
- See example in CC
- Multi-equation stationary process:
  - As we have seen: a system with two variables and first order implies final form equations
  - Those equations are on ARMA-form
  - Hence, by direct reasoning, the theory for ARMA-process in this chapter also covers time series that at are generated by a dynamic system.
  - This brings us to VARs, the topic of Ch. 5.