These lecture notes are totally informal, hence they may contain several mistakes and omissions (of any type).

They do not substitute the lectures or the official notes of the course.

23/2/16

An imprecise introduction to stochastic processes.

What is a random variable?

A random variable is neither random nor a variable. A r.v. is a function from one probability space to another, which preserves the measure characteristics. Thus, its domain is a probability space.

- A probability space is a triple : $(\Omega, \mathcal{F}, \mathbb{P})$

all possible events

domain of r.v.

set of subsets of Ω to which probabilities can be consistentely attributed to (σ-algebra)

probability measure (or distribution) $\mathbb{P}: \mathcal{F} \to \mathbb{R}$ real valued function with some properties

- Consider a measurable space $(\mathbb{R}, \mathcal{B})$

range of r.v.

Borel algebra on $\mathbb{R}$. it is a σ-algebra containing among others, familiar subsets of $\mathbb{R}$ such as intervals

A random variable is a function $x: \Omega \to \mathbb{R}$ that respects the measurable structures (the two σ-algebras)

For any measurable subset of $\mathbb{R}$ i.e. $\forall A \in \mathcal{B}$, its image $x^{-1}(A) \in \mathcal{F}$.

Whatever can be measured in the range, it can be measured through the function x to the domain.

Therefore,

- $(\Omega, \mathcal{F}, \mathbb{P}) + x \to$ essentially defines a probability measure in the set of real numbers.

$\mathbb{P}^*(A) := \mathbb{P}(x^{-1}(A)) \quad \forall A \in \mathcal{B}$

probability distributions that x follows

- What is a stochastic process?

Suppose that we are given $\Theta \neq \emptyset$.

An $\mathbb{R}$-valued stochastic process over Θ is a collection of random variables indexed by θ i.e. $\{x_\theta : \Omega \to \mathbb{R}, \ \theta \in \Theta, \ x_\theta$ is a random variable$\}$ as long as some consistency conditions* hold.

For example, remember the likelihood function

$\forall \mu$ the likelihood $\ell_T(\mu) = -\frac{1}{2}\sum_{i=1}^{T}(y_i - \mu)^2$ is a stochastic process.

$\mu \in \Theta \subseteq \mathbb{R}$

This function can also perceived as a collection of functions from $\Theta \to \mathbb{R}$ since for every value of the sample $(y_i)_{i=1}^{T}$ we can get a collection of fncs $\Theta \to \mathbb{R}$. ⇒ sample path

$\forall \omega$ ⇒ we get function from Θ to $\mathbb{R}$.

Thus,

- Dually, the process $x := \{x_\theta : \Omega \to \mathbb{R}, \ \theta \in \Theta, \ x_\theta$ is a r.v.$\} : \Omega \to$ set of functions $\Theta \to \mathbb{R}$ — range on function space

This must be sth like a random variable, it must respect the measurable structures i.e. the σ-algebras

using the probability distr. here, we define the prob. distr. in function spaces

- A random variable is a stochastic process and so is a random vector

↳ The consistency conditions enables us to study a stochastic process through a joint probability distribution of vectors.

* Daniell - Kolmogorov Extension Thm

Definition: Suppose that x is an $\mathbb{R}$-valued stochastic process over $\Theta \neq \emptyset$
Suppose that $\underset{\neq \emptyset}{\Theta^*} \subseteq \Theta$ that is finite.

$$\Theta^* = \{\theta_1, \theta_2, \dots, \theta_K\}$$

Consider $x_{\theta_1}, x_{\theta_2}, \dots, x_{\theta_K}$ i.e. a finite subset of the stochastic process. The joint distribution of this finite portion of the process x, is termed as a <u>finite dimensional distribution of x (fidi)</u>.

<u>Theorem</u>

The Daniell-Kolmogorov Extension Thm implies that x can be equivalently described by its set of fidis which is equivalent to knowing the probability measure on the function space of the stochastic process.

<u>Example:</u> A <u>Gaussian $\mathbb{R}$-valued process</u> over Θ is a stochastic process in which every fidi is a Normal distribution.

We can prove that a Gaussian process is equivalently described by the following two functions:

(i) the mean fnc $\mu: \Theta \to \mathbb{R}$ where $\mu(\theta) = E(x_\theta)$

(ii) the covariance fnc $\Gamma: \Theta \times \Theta \to \mathbb{R}$ where $\Gamma(\theta, \theta^*) = E(x_\theta x_{\theta^*})$

For <u>example</u>: $z \sim \mathcal{N}(0,1)$, $\Theta = \mathbb{R}$, $x_\theta = \theta \cdot z$ can be proven to be a Gaussian process with $\mu(\theta) = E(x_\theta) = \theta E(z) = 0 \quad \forall \theta \in \mathbb{R}$

and $\Gamma(\theta, \theta^*) = E(x_\theta x_{\theta^*}) = \theta\theta^* E(z^2) = \theta\theta^*$

Definition: An ($\mathbb{R}$-valued) <u>time series</u> is an ($\mathbb{R}$-valued) stochastic process over Θ, for Θ totally ordered (so that it represents a set of time instances)

<u>ex</u> $\mathbb{R}$, $[0,1]$, $\mathbb{Z}$, $\mathbb{N}$

- When Θ is a continuous subset of $\mathbb{R}$ (eg $\mathbb{R}$ or some interval), then we have a time series in continuous time.
When Θ is a discrete subset of $\mathbb{R}$ (eg. $\mathbb{Z}$ or $\mathbb{N}$ or $\mathbb{Q}$), then we have a time series in discrete time.

The previous Gaussian process is a continuous time, time series.
We will usually occupy ourselves with discrete-time, time series over $\mathbb{R}$ or $\mathbb{N}$ (double stochastic sequences and stochastic seq respectively) (except from Brownian motions which are continuous and we are going to study later)

Examples
$\{\dots, x_{-k}, \dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots, x_k, \dots\}$ where the r.v. are iid is an example of a double stochastic sequence : $x = (x_t)_{t \in \mathbb{Z}}$ where the x_t, $t \in \mathbb{Z}$ are iid
... analogously for $x = (x_t)_{t \in \mathbb{N}}$ where x_t, $t \in \mathbb{N}$ are iid, this is an example of a stochastic sequence.

- Suppose that we have a process $x = (x_t)_{t \in \overset{\Theta}{\widehat{\mathbb{Z}}}}$
Consider the finite subset of $\mathbb{Z}$, $\{t_1, t_2, \dots, t_k\}$. This subset defines a random vector $(x_{t_1}, x_{t_2}, \dots, x_{t_k})$ whose distribution describe the probability : $\mathbb{P}(x_{t_1} \in A_1, x_{t_2} \in A_2, \dots, x_{t_k} \in A_k)$ where $A_1, A_2, \dots, A_k \in \mathcal{B}$
{this is a fidi of the process that corresponds to $\{t_1, t_2, \dots, t_k\}$
Given that $\Theta = \mathbb{Z}$, you can translate time just by adding to each element a constant getting again a subset of Θ.
Suppose that $m \in \mathbb{Z}$ and consider the set $\{t_1 + m, t_2 + m, \dots, t_k + m\}$ call this as the m time translation of $\{t_1, t_2, \dots, t_k\}$ and the fidi that corresponds to the m-time translation is
$\mathbb{P}(x_{t_1+m} \in A_1, x_{t_2+m} \in A_2, \dots, x_{t_k+m} \in A_k)$

For a general stochastic process this two fidis won't coincide.

If the two fidis are the same then we say that the fidi remain invariant to time translations.

If $\mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2, \dots, X_{t_k} \in A_k) = \mathbb{P}(X_{t_1+m} \in A_1, X_{t_2+m} \in A_2, \dots, X_{t_k+m} \in A_k)$

$\forall A_1, A_2, \dots, A_k \in \mathcal{B}, \ \forall m \in \mathbb{Z}$

then we say that the particular fidi remains invariant with respect to time translation.

- $X = (X_t)_{t \in \mathbb{Z}}$ is called (strictly) stationary iff every fidi of the process remains invariant of time translations

G Re-write the above definition for $t \in \mathbb{N}$

G Gaussian restricted → not strictly stationary

iid → strictly stationary

2/3/16

- Trying to re-write the definition of stict stationarity for $t \in \mathbb{N}$ the problem we face is the fact that not all time translations are valid in $\mathbb{N}$.

- A typical example of a strictly stationary stochastic sequence, is $x = (x_t)_{t \in \mathbb{Z}}$, x_t are iid.
  In order to prove it, consider an arbitrary fidi.
  Suppose that $t_1, t_2, \ldots, t_k \in \mathbb{Z}$ and an analogous set of measurable events $A_1, A_2, \ldots, A_k$

$$\mathbb{P}(x_{t_1} \in A_1, x_{t_2} \in A_2, \ldots, x_{t_k} \in A_k) \overset{\text{ind.}}{=} \prod_{i=1}^{k} \mathbb{P}(x_{t_i} \in A_i) \overset{\text{homog.}}{=} \prod_{i=1}^{n} \mathbb{P}(x_{t_i+m} \in A_i)$$

$$\overset{\text{ind}}{=} \mathbb{P}(x_{t_1+m} \in A_1, x_{t_2+m} \in A_2, \ldots, x_{t_k+m} \in A_k) \quad \forall m \in \mathbb{Z}$$

Since this fidi is chosen arbitrarily, every fidi is invariant to time translations $\Rightarrow (x_t)_{t \in \mathbb{Z}}$ is strictly stationary

<u>FACT 1.</u> If $x = (x_t)_{t \in \mathbb{Z}}$ is strictly stationary then the marginal distribution of x_t is independent of t.
(every marginal is essentially a fidi of dimension 1 and <u>every</u> fidi is invariant to time translations)

However, the converse does not hold in general.

G Consider a random vector of 3 dimensions and consider it as a stochastic process which follows a joint normal distribution. Find a covariance matrix so that the covariance between the 1^{st} and the 2^{nd} element $\neq$ covariance between the 2^{nd} and the 3^{rd} element.

<u>FACT 2.</u> If $x = (x_t)_{t \in \mathbb{Z}}$ is comprised by (jointly) independent random variables then x is strictly stationary iff the marginals do not depend on t.

Counter example. $z \sim \mathcal{N}(0,1)$. Define $x = (x_t)_{t\in\mathbb{Z}}$ by $x_t := tz \;\; \forall\, t \in \mathbb{Z}$ restricted to the set of integers.

Thus $E(x_t) = 0$, $V(x_t) = t^2 \Rightarrow x_t \sim \mathcal{N}(0, t^2)$

Therefore, the marginal distribution depends on t
& not strictly stationary

Weak (or covariance or second order) Stationarity

Definition: $x = (x_t)_{t\in\mathbb{Z}}$ is weakly stationary iff
i) $E(x_t) \in \mathbb{R}$ and is independent of t
ii) $V(x_t) < +\infty$ and is independent of t
iii) $Cov(x_t, x_{t-k})$ $(k \in \mathbb{Z})$ must be independent of t
(it could depend on k)
(we can show that this covariance is always well-defined using the fact that ii) holds and the Cauchy-Schwarz inequality)

ex The Gaussian process defined in the previous counter example, is neither strictly, nor weakly stationary.

Definition: The process $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ is a white noise process $(WN(\sigma^2))$ iff
i) $E(\epsilon_t) = 0 \;\; \forall\, t \in \mathbb{Z}$
ii) $V(\epsilon_t) = \sigma^2 > 0 \;\; \forall\, t \in \mathbb{Z}$
iii) $Cov(\epsilon_t, \epsilon_{t-k}) = 0 \;\; \forall\, t \in \mathbb{Z}, \; k \neq 0$

- A white noise process is weakly stationary.

- We will use WN as a building block.

Example: iid $\mathcal{N}(0,1)$

<u>Example</u>. $x_t \sim t_2$ iid $\Rightarrow$ 2nd moment does not exist, $E(x_t^2) < +\infty \ \forall t$

therefore x_t is not weakly stationary.

Since it is iid $\Rightarrow$ it is strictly stationary.

(WN $\Rightarrow$ uncorrelatedness but WN $\not\Rightarrow$ independence)

<u>Example</u>. $x = (x_t)_{t \in \mathbb{Z}}$, iid, $x_t \sim t_4 \Rightarrow E(x_t^2) < +\infty \Rightarrow$

x_t is both covariance and strictly stationary.

- $\gamma_k := Cov(x_t, x_{t-k})$ as a function of k.

  autocovariance fnc of the process

<u>Definition</u>: A weakly stationary process is called <u>regular</u> iff

$\gamma_k \to 0$ as $k \to \infty$.

- Regularity implies an asymptotic uncorrelatedness.

<u>Definition</u>: If $\sum_{k=0}^{\infty} |\gamma_k| < +\infty$ then the weakly stationary process is called <u>short memory</u>.

- $\sum_{k=0}^{\infty} |\gamma_k| < +\infty \Rightarrow \gamma_k \to 0$ as $k \to \infty$ thus short memory implies regularity. However, the converse does not hold (<u>ex</u>. if $\gamma_k = \frac{1}{k+1}$ even if $\gamma_k \to 0$ as $k \to \infty$, $\sum_{k=0}^{\infty} \frac{1}{k+1} = \infty$)

- Short memory implies that elements of the process become fast enough uncorrelated asymptotically.

- $\rho_k := \frac{\gamma_k}{\gamma_0}$ autocorrelation function of a weakly stationary process.

Linear (causal) processes

Suppose that $\epsilon = (\epsilon_t)_{t\in \mathbb{Z}}$ is a strictly stationary white noise process. Suppose also that we are given a real sequence $(b_j)_{j\in\mathbb{N}}$ which is *absolutely summable that is $\sum_{j=0}^{\infty} |b_j| < +\infty$.

(e.g. $b_j = b^j$, $|b|<1$,
$$\sum_{j=0}^{\infty}|b_j| = \sum_{j=0}^{\infty}|b^j| = \frac{1}{1-|b|} < +\infty \Big)$$

<u>Definition</u>: The <u>linear process</u> $y = (y_t)_{t\in\mathbb{Z}}$ defined upon the process ϵ and the sequence of coefficients $(b_j)_{j\in\mathbb{N}}$ is defined by
$$y_t := \sum_{j=0}^{\infty} b_j \epsilon_{t-j} \quad \forall\, t\in\mathbb{Z}$$
It is a well defined process being the limit of convergence in mean of $\sum_{j=0}^{n} b_j \epsilon_{t-j}$ as $n\to\infty$.
(We can prove it using the fact that is a limit + dominated convergence th^m ⇒ expectations go in the infinite sum)

<u>Properties</u>:

1. y is strictly stationary

2. y is weakly stationary

i) $E(y_t) = E\left(\sum_{j=0}^{\infty} b_j \epsilon_{t-j}\right) = \sum_{j=0}^{\infty} E(b_j\epsilon_{t-j}) = \sum_{j=0}^{\infty} b_j \underbrace{E(\epsilon_{t-j})}_{=0} = 0 \quad \forall\, t\in\mathbb{Z}$

since ϵ is WN

ii), iii) $\forall\, k \geq 0$
$$E(y_t y_{t-k}) = E\left(\sum_{j=0}^{\infty} b_j\epsilon_{t-j} \sum_{j=0}^{\infty} b_j \epsilon_{t-k-j}\right) = E\left(\sum_{j=0}^{\infty}\sum_{i=0}^{\infty} b_j\epsilon_{t-j}\, b_i \epsilon_{t-k-i}\right)$$
$$= E\left(\sum_{i,j=0}^{\infty} b_j b_i \epsilon_{t-j}\epsilon_{t-k-i}\right) = \sum_{i,j=0}^{\infty} E(b_j b_i \epsilon_{t-j}\epsilon_{t-k-i})$$

* a weaker condition is square summability

$$= \sum_{i,j=0}^{\infty} b_j b_i \, E(\epsilon_{t-j} \, \epsilon_{t-k-i})$$

but $E(\epsilon_{t-j}, \epsilon_{t-k-i}) = \begin{cases} E(\epsilon_0^2) & \text{iff } j = k+i \\ 0 & \text{iff } j \neq k+i \end{cases}$

thus $E(y_t \, y_{t-k}) = \sum_{\substack{i,j=0 \\ j=k+i}}^{\infty} b_j b_i \, E(\epsilon_0^2) + \sum_{\substack{i,j=0 \\ j\neq k+i}}^{\infty} b_j b_i \cdot 0 = \sum_{\substack{i,j=0 \\ j=k+i}}^{\infty} b_j b_i \, E(\epsilon_0^2)$

$$= \sum_{i=0}^{\infty} b_{k+i} \, b_i \, E(\epsilon_0^2) = E(\epsilon_0^2) \sum_{i=0}^{\infty} b_{k+i} \, b_i$$

We can prove that $\sum_{i=0}^{\infty} b_{k+i} b_i$ is a real number if $\sum_{i=0}^{\infty} |b_{k+i} b_i|$ converges, which is equivalent to $\sum_{i=0}^{\infty} |b_{k+i} b_i| < +\infty$.

We can also prove that $\sum_{i=0}^{\infty} |b_{k+i} b_i| \leq \sqrt{\sum_{i=0}^{\infty} b_{k+i}^2} \sqrt{\sum_{i=0}^{\infty} b_i^2}$

$$\leq \sqrt{\sum_{i=0}^{\infty} |b_{k+i}|} \sqrt{\sum_{i=0}^{\infty} |b_i|} \qquad (*)$$

Also, $\sqrt{\sum_{i=0}^{\infty} |b_{k+i}|} \leq \sqrt{\sum_{i=0}^{\infty} |b_i|}$

so $(*) \leq \left(\sqrt{\sum_{i=0}^{\infty} |b_i|} \right)^2 = \sum_{i=0}^{\infty} |b_i| < +\infty$

Therefore, $E(y_t \, y_{t-k})$ exists for every t and every k $\Rightarrow$ conditions ii) and iii) hold.

8/3/16

Linear Causal Processes

- $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ strictly stationary WN $(E\epsilon_0^2)$
- $(b_j)_{j \in \mathbb{N}}$ absolute summable $\left[\sum_{j=0}^{\infty} |b_j| < +\infty\right]$

$$y := (y_t)_{t \in \mathbb{Z}}, \quad y_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j}$$

↳ Absolute summability along with strictly stationary implies that y is strictly stationary.

↳ WN along with absolute summability implies that y is weakly stationary

- $E y_t = 0 \ \forall \ t \in \mathbb{Z}$
- $E y_t y_{t-k} = E(\epsilon_0^2) \sum_{j=0}^{\infty} b_j b_{j+k}$ which is a real number $\forall \ k \in \mathbb{Z}$.

- $\rho_k = \dfrac{\gamma_k}{\gamma_0} = \dfrac{\sum_{k=0}^{\infty} b_j b_{j+k}}{\sum_{j=0}^{\infty} b_j^2}$ • $\gamma_k = E(\epsilon_0^2) \sum_{j=0}^{\infty} b_j b_{j+k}$

Example: Suppose that $b_j = b^j, \ |b| < 1, \ j \in \mathbb{N}$

This is an example of an AR(1) process.

$$\sum_{j=0}^{\infty} |b^j| = \sum_{j=0}^{\infty} |b|^j = \frac{1}{1-|b|} < +\infty$$

All the above hold for this example $\Rightarrow$ stict stationarity
weak stationarity

$$\gamma_k = \sum_{j=0}^{\infty} b_j b_{j+k} = \sum_{j=0}^{\infty} b^j b^{j+k} = b^k \sum_{j=0}^{\infty} b^{2j} = b^k \sum_{j=0}^{\infty} (b^2)^j = \frac{b^k}{1-b^2}$$

As $k \to \infty$, $\gamma_k \to 0$ i.e. in this example absolute summability implies regularity.

$$\sum_{k=0}^{\infty} |\gamma_k| = \sum_{k=0}^{\infty} \left| \frac{\theta^k}{1-\theta^2} \right| = \sum_{k=0}^{\infty} \frac{|\theta^k|}{1-\theta^2} = \frac{1}{1-\theta^2} \sum_{k=0}^{\infty} |\theta|^k$$

$$= \frac{1}{1-\theta^2} \frac{1}{1-|\theta|^k} < +\infty \qquad \text{short memory}$$

↳ Short memory implies sth about the speed wrt which the autocovariances goes to 0.

$\rho_k = \theta^k$

Strict stationarity and Transformations

Suppose that $x = (x_t)_{t \in \mathbb{Z}}$ and $y = (y_t)_{t \in \mathbb{Z}}$ are strictly stationary processes. We question whether strict stationarity remains invariant over certain transformations.
We can define a new process that is the sum of those two process. The resulting random variable for each t results from the pointwise addition of the two processes.

FACT 1. $x + y := (x_t + y_t)_{t \in \mathbb{Z}}$ is a well defined stochastic process over $\mathbb{Z}$. If the additive processes are strictly stationary then the sum is also strictly stationary.

FACT 2. Given a process and λ a real number, $\lambda \in \mathbb{R}$, then $\lambda x := (\lambda x_t)_{t \in \mathbb{Z}}$ is well defined and strictly stationary. Thus, strict stationarity is preserved through scalar multiplication

↳ Obviously, we can combine those two facts

FACT 3. Suppose $f: \mathbb{R} \to \mathbb{R}$ is appropriately measurable. Then $f(x_t)$ is also a random variable, so we can construct a well defined

stochastic process $(f(x_t))_{t\in\mathbb{Z}}$ by pointwise composition.
Thus, if $(x_t)_{t\in\mathbb{Z}}$ is strictly stationary then $(f(x_t))_{t\in\mathbb{Z}}$ is also strictly stationary.

e.g. $f(x) = x^2$ along with the previous AR(1) process
$$\left(\left(\sum_{j=0}^{\infty} \beta^j \epsilon_{t-j}\right)^2\right)_{t\in\mathbb{Z}}$$

FACT 2*. Define a new stochastic process by pointwise multiplication of two strictly stationary stoch. processes
$x \cdot y := (x_t \cdot y_t)_{t\in\mathbb{Z}}$
If (x_t) and (y_t) are strictly stationary then $(x_t y_t)_{t\in\mathbb{Z}}$ is also str. stationary.

↳ However, weak stationarity may be lost under such transformations because weak stationarity implies restrictions on moments.

FACT 4. (L is the lag operator)
If we define $Lx := (Lx_t)_{t\in\mathbb{Z}} = (x_{t-1})_{t\in\mathbb{Z}}$ then it is strictly stationary.

e.g. $(y_t \epsilon_{t-1})_{t\in\mathbb{Z}}$
$y = (y_t)_{t\in\mathbb{Z}}$ is strictly stationary
$\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ is a WN thus strictly stationary
$\Downarrow 4$
$L\epsilon = (L\epsilon_t)_{t\in\mathbb{Z}} = (\epsilon_{t-1})_{t\in\mathbb{Z}}$ is strictly stationary
$\Downarrow 2^*$
$y \cdot L\epsilon = (y_t \epsilon_{t-1})_{t\in\mathbb{Z}}$ is strictly stationary

FACT 4⁺. The same holds for weak stationarity
i. $Ex_{t-1} = Ex_t \in \mathbb{R}$, independent of t, due to weak stat. of x
ii. $Vx_{t-1} = Vx_t < +\infty$, —"— —"— —"— —"— —"— —"—
iii. $Cov(x_{t-1}, x_{t-k-1}) = Cov(x_t, x_{t-k})$, —"— —"— —"— —"— —"—

G Prove FACT 2 for weak stationarity
G Prove that weak stat remains invariant to transformations defined by polynomials.

Impresice discussion about ergodicity

Doob's LLN : If $x = (x_t)_{t \in \mathbb{Z}}$ is strictly stationary and $E|x_0| < +\infty$ ($\Leftrightarrow E(x_0) \in \mathbb{R}$) then $\frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow[T \to \infty]{} E(x_0 / \mathcal{J})$ a.s.

Note that, this is actually a random variable.

$\mathcal{J}$ information set that contains information about properties of L when applied to the process

G Are there any conditions under which this conditional expectation is trivial?
Ergodicity is a necessary and sufficient condition.
Thus, ergodicity is equivalent to, that $\mathcal{J}$ is "trivial" in the sense that $E(x_0/\mathcal{J}) = E(x_0)$.
Ergodicity is also equivalent to that, if $f, g : \mathbb{R} \to \mathbb{R}$ are appropriately measurable, $Cov(f(x_t), g(x_{t-k}))$ when well defined, then as $k \to \infty$ $Cov(f(x_t), g(x_{t-k})) \to 0$.
(asymptotic independence).

A typical example of an ergodic process is an iid process.

FACT 5. FACT 1-4 hold also for ergodicity.

G Is an ergodic process a regular process?

Asymptotic independence $\Rightarrow$ asymptotic uncorrelatedness
(ex f, g : identity fncs, and assume that the second moments exist)

FACT 6. Suppose that $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process. For some bivariate real function $f(x, y)$, consider a stochastic difference equation or stochastic recursion, $y_t = f(\epsilon_t, y_{t-1})$.

If f is appropriately measurable the derivative of f wrt the second argument i.e. y exists for every x and is also an appropriately measurable function, some other conditions hold, <u>if</u> $E\Big(\underbrace{\ln\Big(\sup_y \Big|\frac{\partial f}{\partial y}(\epsilon, y)\Big|\Big)}_{\text{Lipschitz coefficient}}\Big) < 0$ (or $-\infty$)

<u>then the recurrence equation defined above has a unique stationary and ergodic solution</u> obtained as an appropriate limit of backward substitution.

$$y_t = f(\epsilon_t, y_{t-1}) = f\big(\epsilon_t, f(\epsilon_{t-1}, y_{t-2})\big) = f\Big(\epsilon_t, f\big(\epsilon_{t-1}, f(\epsilon_{t-2}, y_{t-3})\big)\Big)$$
$$= \dots$$

The above says that we obtain an ergodic and stationary solution as the limit of this procedure.

<u>Example:</u> AR(1) Recursion. $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic $WN(E(\epsilon_0^2))$.

$$y_t = b y_{t-1} + \epsilon_t, \quad y_t = f(\epsilon_t, y_{t-1})$$

$$f(x, y) = x + by, \quad \left|\frac{\partial f(x,y)}{\partial y}\right| = |b| \quad \forall x$$

$$\sup_y \left|\frac{\partial f(x,y)}{\partial y}\right| = |b|$$

hence, $\ln\Big(\sup_y \Big|\frac{\partial f(x,y)}{\partial y}\Big|\Big) = \ln|b|$

Thus, $E\Big(\ln\Big(\sup_y \Big|\frac{\partial f(x,y)}{\partial y}\Big|\Big)\Big) = E(\ln|b|) = \ln|b|$ which is less than 0 or $-\infty$ when $|b| < 1$

$\Rightarrow$ there exists a unique stationary ergodic solution given as:

$$y_t = \epsilon_t + b y_{t-1} = \epsilon_t + b\epsilon_{t-1} + b^2 y_{t-2} = \epsilon_t + b\epsilon_{t-1} + b^2 \epsilon_{t-2} + b^3 y_{t-3}$$

$$= \epsilon_t + b\epsilon_{t-1} + b^2 \epsilon_{t-2} + b^3 \epsilon_{t-3} + b^4 y_{t-4} \dots$$

$$= \sum_{j=0}^{\infty} b^j \epsilon_{t-j}$$

Birkoff's LLN

If $x = (x_t)_{t \in \mathbb{Z}}$ is strictly stationary and ergodic and $E|x_0| < +\infty$ then

$$\frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow[T \to \infty]{} E(x_0) \quad \text{a.s.}$$

G. In the context of a stationary ergodic AR(1) consider

$$\frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \qquad \text{(OLSE for } b \text{ in the AR(1) context)}$$

So what is the statistical model?

$Y = Xb + \epsilon$

where $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}$, $X = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{T-1} \end{pmatrix}$, $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}$

- rank condition (fcr of X)
- strict exogeneity implies uncorrelatedness of every element of ϵ for every regression. But it is also obvious that y_1 is correlated to ϵ_1 (remember that $y_1 = \epsilon_1 + \dots$)

  $\Rightarrow$ unbiased.

15/3/16

In the context of AR(1) we were occupied with:

▸ $b_T = \dfrac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$

- Strict exogeneity does not hold $\Rightarrow$ biased estimator
- Horizontal uncorrelatedness $\Rightarrow$ weak exogeneity which implies in our context consistency

Q: What if not $\Theta = \mathbb{R}^n$?

We can show that asymptotically if $b_0 \in \operatorname{Int}\Theta$ then we have good properties.

$$b_T = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \frac{\sum_{t=1}^T (b y_{t-1} + \epsilon_t) y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \frac{\sum (b y_{t-1}^2 + \epsilon_t y_{t-1})}{\sum y_{t-1}^2} = \frac{b \sum y_{t-1}^2 + \sum \epsilon_t y_{t-1}}{\sum y_{t-1}^2}$$

$$= b + \frac{\sum \epsilon_t y_{t-1}}{\sum y_{t-1}^2} = b + \frac{\frac{1}{T}\sum \epsilon_t y_{t-1}}{\frac{1}{T}\sum y_{t-1}^2}$$

- $(\epsilon_t y_{t-1})_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic since (ϵ_t) and (y_t) are stationary ergodic.
- $E(\epsilon_0 y_{-1}) = E\left(\epsilon_0 \sum_{i=0}^{\infty} b^i \epsilon_{-1-i}\right) = E\left(\sum_{i=0}^{\infty} b^i \epsilon_0 \epsilon_{-1-i}\right) = \sum_{i=0}^{\infty} b^i E(\epsilon_0 \epsilon_{-1-i}) = 0$

since (ϵ_t) is WN $\Rightarrow E(\epsilon_0 \epsilon_{-1-i}) = 0 \ \forall i$

▸ From Birkoff's LLN:

$\frac{1}{T}\sum_{t=1}^T \epsilon_t y_{t-1} \longrightarrow 0$ a.s.

- $(y_{t-1})_{t\in\mathbb{Z}}$ is stationary ergodic since (y_t) is stationary ergodic

  & $(y_{t-1}^2)_{t\in\mathbb{Z}}$ is also stationary ergodic
- $E(y_0^2) = \dfrac{E(\epsilon_0^2)}{1-b^2}$

→

▸ From Birkoff's LLN:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \frac{E(\epsilon_0^2)}{1-\theta^2} \neq 0 \quad \text{a.s.}$$

(It could be 0 if $E(\epsilon_0^2)=0$ but this would mean that $E(\epsilon_0)=0$, $E(\epsilon_0^2)=0$, the distribution of ϵ would be degenerate)

· Due to the fact that 0 and $\frac{E(\epsilon_0^2)}{1-\theta^2} \in \mathbb{R}$, even we have established explicitly each limit, since these limits are degenerate random variables thus constants, we can establish the joint limit of them:

$$\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \epsilon_t y_{t-1} \\ \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 \\ \frac{E(\epsilon_0^2)}{1-\theta^2} \end{pmatrix} \quad \text{a.s.}$$

What about the quotient, does it converge to the quotient of the limits?

Since $\frac{E(\epsilon_0^2)}{1-\theta^2} \neq 0$ due to CMT:

$$\frac{\frac{1}{T}\sum_{t=1}^{T} \epsilon_t y_{t-1}}{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \frac{0}{\frac{E(\epsilon_0^2)}{1-\theta^2}} = 0$$

Hence, $\theta_T \to \theta + 0$ a.s. i.e. OLS is strongly consistent

Now, in order to establish a CLT in the context of stationary ergodic stochastic processes, we need to establish the rate of which the autocovariance goes to zero.

Martingal Difference processes

Definition: A *filtration* on $(\Omega, \mathcal{F}, P)$, is a (double-) sequence $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ where $\mathcal{F}_t \subseteq \mathcal{F}$ and also a σ-algebra $\forall t \in \mathbb{Z}$, that is monotonic.

i.e. $s \le t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$

think of it as a well defined information set

↳ a sequence of information sets that does not diminish in time

Examples

1. $f_t = f \quad \forall\, t \in \mathbb{Z}$ constant
2. Given a process $(x_t)_{t\in\mathbb{Z}}$, $f_t = \sigma(x_{t-i}, i \geq 0)$

this filtration follows the sequence of the process

Definition: Given a filtration $(f_t)_{t\in\mathbb{Z}}$ and a process $(x_t)_{t\in\mathbb{Z}}$, we say that the process is adopted to the filtration iff x_t is measurable wrt f_t, $\forall\, t \in \mathbb{Z}$

Examples

1. $f_t = f \quad \forall\, t \in \mathbb{Z}$ every well defined process is adopted to this filtration.
2. $f_t = \sigma(x_{t-i}, i \geq 0)$ by construction $(x_t)_{t\in\mathbb{Z}}$ is adopted to $(f_t)_{t\in\mathbb{Z}}$.

Definition: The pair $((x_t)_{t\in\mathbb{Z}}, (f_t)_{t\in\mathbb{Z}})$ is a martingale difference process (or equivallently the process $(x_t)_{t\in\mathbb{Z}}$ is a martingale difference process wrt the filtration $(f_t)_{t\in\mathbb{Z}}$) iff

1. $(x_t)_{t\in\mathbb{Z}}$ is adopted to the filtration $(f_t)_{t\in\mathbb{Z}}$.
2. $E|x_t| < +\infty \quad \forall\, t \in \mathbb{Z}$
3. $E(x_t / f_{t-1}) = 0 \quad \forall\, t \in \mathbb{Z}$

(does not require anything as stationarity and ergodicity)

↳ Conditional expectations are well defined if unconditional expectations exist. Thus, 2. implies that conditional expectations appearing in 3. are well defined.

↳ $E(x_t) \overset{LIE}{=} E(E(x_t / f_{t-1})) = E(0) = 0 \quad \forall\, t \in \mathbb{Z}$

↳ Let $f: \mathbb{R} \to \mathbb{R}$ measurable, the $f(x_{t-k})$, $k>0$ is a well-defined r.v.
If $E(x_t f(x_{t-k})) \in \mathbb{R}$ then we can evaluate it using LIE since the conditional expectation is well-defined.

$$E(x_t f(x_{t-k})) \overset{LIE}{=} E\Big(E\big(x_t f(x_{t-k}) / f_{t-1}\big)\Big) \overset{1.}{=} E\Big(f(x_{t-k})\, E(x_t / f_{t-1})\Big)$$
$$\overset{3.}{=} E\big(f(x_{t-k}) \cdot 0\big) = 0$$

due to 1. $f(x_{t-k})$ belongs somehow to f_{t-1} ⇒ it can come out of the expectation.

Thus,
For an m.d. the covariances are well-defined.

Suppose that f is the identity function, then if the above expectation exists, the elements of the process are uncorrelated.
Also, $|E(x_t x_{t-k})| \leq \sqrt{E(x_t^2)}\sqrt{E(x_{t-k}^2)}$ existing 2nd moments ⇒ uncorrelatedness.
When $E(x_t^2) < +\infty \;\; \forall\, t \in \mathbb{Z} \xRightarrow{\text{Cauchy-Schwarz}} E(x_t x_{t-k}) \in \mathbb{Z} \;\; \forall\, k>0$
hence $Cov(x_t, x_{t-k}) = 0 \;\; \forall\, t \in \mathbb{Z},\; \forall\, k>0$

<u>Definition:</u> The pair $((x_t)_{t\in\mathbb{Z}}, (f_t)_{t\in\mathbb{Z}})$ is a <u>square-integrable martingale difference</u> iff
1. $(x_t)_{t\in\mathbb{Z}}$ is adopted to $(f_t)_{t\in\mathbb{Z}}$
2. $E(x_t^2) < +\infty, \;\; \forall\, t \in \mathbb{Z}$
3. $E(x_t / f_{t-1}) = 0, \;\; \forall\, t \in \mathbb{Z}$

- An sMd implies uncorrelatedness
- Is an sMd a WN? Not necessarily, we don't know if moments are equal across t, they may depend on t.
- Is a strict stationary and sMd process, a WN? Strict stationarity implies that the existing second moments are equal, so yes.

Corrolary. If $(x_t)_{t\in\mathbb{Z}}$ is strictly stationary and sMd process wrt some filtration and $E(x_0^2) > 0$, then it is a WN process.

(Since it is an m.d $\Rightarrow$ mean $= 0$ and since it is an s.M.d $\Rightarrow$ Cov $= 0$.
Str. Stationarity + s.Md $\Rightarrow$ 2nd moments exist and are independent of t, $\forall t \in \mathbb{Z}$. $\Rightarrow$ W.N. by construction)

Example: Suppose that $(z_t)_{t\in\mathbb{Z}}$ iid, and $E(z_0)=0$. For $R \geq 0$ consider the process $(\epsilon_t)_{t\in\mathbb{Z}}$, $\epsilon_t := \prod_{j=0}^{R} z_{t-j}$ $\forall t \in \mathbb{Z}$.

Since $(z_t)_{t\in\mathbb{Z}}$ are iid, it is a strictly stationary and ergodic process.
Since (ϵ_t) is the pointwise product of stationary ergodic processes, it is also stationary ergodic.
Consider the filtration $f_t = \sigma(z_{t-i}, i \geq 0)$ $\forall t \in \mathbb{Z}$

1. ϵ_t is obtained to f_t
2. $E|\epsilon_t| = E\left|\prod_{j=0}^{R} z_{t-j}\right| = E\left(\prod_{j=0}^{R} |z_{t-j}|\right) \overset{\text{indep.}}{=} \prod_{j=0}^{R} E(|z_{t-j}|) \overset{\text{homog.}}{=} \prod_{j=0}^{R}(E|z_t|)$

$$= (E|z_t|)^{R+1} < +\infty \quad \text{which exists since } E(\epsilon_0) \text{ exists.}$$

3. $E(\epsilon_t / f_{t-1}) = E\left(\prod_{j=0}^{R} z_{t-j} / f_{t-1}\right) = E\left(z_t \prod_{j=1}^{R} z_{t-j} / f_{t-1}\right)$

$$= \prod_{j=1}^{R} z_{t-j}\, E(z_t / f_{t-1}) \overset{\text{indep.}}{=} \prod_{j=1}^{R} z_{t-j}\, E(z_t) = 0 \quad \forall t \in \mathbb{Z}$$

Therefore, $(\epsilon_t)_{t\in\mathbb{Z}}$ is an m.d. process wrt. $(f_t)_{t\in\mathbb{Z}}$ where $f_t = \sigma(z_{t-i}, i \geq 0)$

Q. If we consider the filtration $f_t = \sigma(\epsilon_{t-i}, i \geq 0)$, would the above arguments hold?

- A filtration obtained by ϵ_t is smaller than a filtration obtained by the original z_t, since it contains products of the original.

What about sMd?

We need to assume sth about the second moments.

Suppose now that $E(z_0^2) < +\infty$.

Then, $E(\epsilon_t^2) = E\left(\prod_{j=0}^{R} z_{t-j}\right)^2 = E\left(\prod_{j=0}^{R} z_{t-j}^2\right) \overset{\text{indep.}}{=} \prod_{j=0}^{R} E(z_{t-j}^2) \overset{\text{homog.}}{=} \prod_{j=0}^{R} E(z_t^2)$

$$= \left(E(z_0^2)\right)^{R+1} < +\infty$$

Therefore, $(\epsilon_t)_{t\in\mathbb{Z}}$ is an sMd wrt $(f_t)_{t\in\mathbb{Z}}$ where $f_t = \sigma(z_{t-i}, i \geq 0)$

$\Rightarrow$ this is a WN.

$$V(\epsilon_t / f_{t-1}) = E(\epsilon_t^2 / f_{t-1}) = E\left(\left(\prod_{j=0}^{R} z_{t-j}\right)^2 / f_{t-1}\right) = E\left(\prod_{j=0}^{R} z_{t-j}^2 / f_{t-1}\right)$$

$$= E\left(z_t^2 \prod_{j=1}^{R} z_{t-j}^2 / f_{t-1}\right) \overset{\text{measurable}}{=} \prod_{j=1}^{R} z_{t-j}^2 \cdot E(z_t^2 / f_{t-1})$$

$$= \prod_{j=1}^{R} z_{t-j}^2 \, E(z_t^2) = E(z_0^2) \prod_{j=1}^{R} z_{t-j}^2$$ which is a random variable that depends on t.

Hence, this is an example of a process that is conditionally heteroskedastic if $R > 0$, in the sence that its conditional variance is stochastic and time dependent.

(If $R=0$, (ϵ_t) is iid so we can't have conditional heteroskedasticity)

22/3/15

Example. $(z_t)_{t\in\mathbb{Z}}$ iid, $E(z_0)=0$, $0 < E(z_0^2) < +\infty$

Define $(\epsilon_t)_{t\in\mathbb{Z}}$, $\epsilon_t := \prod_{j=0}^{R} z_{t-j}$

Given the existence of the 2nd moments and the iidness, we proved that (ϵ_t) is an s.M.d. process wrt $(f_t)_{t\in\mathbb{Z}}$, $f_t = \sigma(z_{t-i}, i \geq 0)$ where the filtration is defined on the history of z.

- $E(\epsilon_t)=0$
- $E(\epsilon_t^2) = \left[E(z_0^2)\right]^{R+1}$

Also, we know that given the way this process is constructed, it is a stationary ergodic process. Moreover, it exchibits conditional heteroskedastic when $R>0$, $V(\epsilon_t / f_{t-1}) = E(z_0^2)\prod_{j=1}^{R} z_{t-j}^2$ which is stochastic and depends on t.

CLT for s.M.d. process

Theorem. When $(x_t)_{t\in\mathbb{Z}}$ is a stationary ergodic s.M.d (wrt some filtration) then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \xrightarrow{d} z \sim N(0, E(x_0^2))$

Example. Given that (ϵ_t) is defined as above

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t \xrightarrow{d} z \sim N\left(0, \left[E(z_0^2)\right]^{R+1}\right)$$

Application. The rate and limit distribution of the OLSE in the context of a stationary ergodic AR(1) ($\Rightarrow$ which implies that ϵ_t is WN and $|b|<1$)

$$b_T = b + \frac{1/T \sum_{t=1}^{T}\epsilon_t x_{t-1}}{1/T\sum_{t=1}^{T} x_{t-1}^2} \Leftrightarrow b_T - b = \underbrace{\frac{1}{1/T\sum_{t=1}^{T} x_{t-1}^2}}_{\text{sth positive}} \underbrace{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t x_{t-1}}_{\text{the limit distribution is going to be determined by this part appropriately scaled in order to apply a CLT for sMd} \Rightarrow \text{therefore rescale } b_T - b} \Leftrightarrow$$

$$\sqrt{T}\,(\beta_T - \beta) = \frac{1}{1/T \sum_{t=1}^{T} x_{t-1}^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t x_{t-1}$$

The rate of the estimator is $\sqrt{T}$.

We are examing the properties of the $(\epsilon_t x_{t-1})_{t\in\mathbb{Z}}$ process and we have already proved that is stationary ergodic.

At first let's assume that ϵ_t is sMd w.r.t each natural filtration consisting of its history.

__Assumption__: $(\epsilon_t)_{t\in\mathbb{Z}}$ is an sMd process wrt $(f_t)_{t\in\mathbb{Z}}$, $f_t := \sigma(\epsilon_{t-i}, i \geq 0)$

But do we really need to specify the filtration? We can say that (ϵ_t) is an sMd wrt some filtration which is a weaker condition.

So,

__Assumption*__: $(\epsilon_t)_{t\in\mathbb{Z}}$ is an sMd wrt some filtration.

Is this assumption sufficient?

If we denote the previous filtration with $(f_t)_{t\in\mathbb{Z}}$ we have:

- Remember that, $x_{t-1} = \sum_{j=0}^{\infty} \beta^j \epsilon_{t-j-1}$

$$\text{then} \quad \epsilon_t x_{t-1} = \epsilon_t \sum_{j=0}^{\infty} \beta^j \epsilon_{t-j-1} = \sum_{j=0}^{\infty} \beta^j \epsilon_t \epsilon_{t-j-1}$$

So if (ϵ_t) is adopted to f_t $\xRightarrow[\text{measurability of } \epsilon_t \text{ wrt } f_t]{\text{monotonicity of } f_t}$

(ϵ_t is measurable wrt f_t, then ϵ_{t-1} is also measurable wrt f_t and so does their product $\Rightarrow$) $(\epsilon_t x_{t-1})$ is adopted to f_t (to the same filtration wrt which the (ϵ_t) process is adopted)

- We already evaluated $E(\epsilon_t x_{t-1}) = 0$ (using the fact that ϵ_t is stationary ergodic process) $\Rightarrow$ the 1st moment exist.

- $E(\epsilon_t x_{t-1} / f_{t-1}) = ?$

(ϵ_t) is adopted to f_{t-1} and $(x_t) = \sum_{j=0}^{\infty} \beta^j \epsilon_{t-j}$ is also measurable wrt f_{t-1} $\Rightarrow$

$E(\epsilon_t x_{t-1} / f_{t-1}) = x_{t-1} E(\epsilon_t / f_{t-1}) = 0 \quad \forall\, t \in \mathbb{Z}$

So we proved that the Assumption* (+ stationarity and ergodicity of (ϵ_t)) imply that $(\epsilon_t x_{t-1})_{t\in\mathbb{Z}}$ is an md process wrt the same filtration.

Q: If we don't assume that ϵ_t is stationary ergodic (which along with sMd implies that ϵ_t is a WN), does the last hold?

- Let's see the 4th condition.

$$E(\epsilon_t x_{t-1})^2 = E(\epsilon_t^2 x_{t-1}^2) = E\left[\epsilon_t \sum_{j=0}^{\infty} b^j \epsilon_{t-j-1}\right]^2$$

$$= E\left[\epsilon_t^2 \left(\sum_{j=0}^{\infty} b^{2j} \epsilon_{t-j-1}^2 + \sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} b^i b^j \epsilon_{t-j-1}\epsilon_{t-i-1}\right)\right]$$

$$= \underbrace{\sum_{j=0}^{\infty} b^{2j} E(\epsilon_t^2 \epsilon_{t-j-1}^2)}_{A} + \underbrace{\sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} b^i b^j E(\epsilon_t^2 \epsilon_{t-j-1}\epsilon_{t-i-1})}_{B}$$

no already stated assumptions specify that those sums converge. If A and B exist (in the sense that they converge as real series) then $E(\epsilon_t x_{t-1})^2$ exists and is independent of t due to stationarity.

So,

Assumption: $\sum_{j=0}^{\infty} b^{2j} E(\epsilon_t^2 \epsilon_{t-j-1}^2) < +\infty$ and $\left| \sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} b^j b^i E(\epsilon_t^2 \epsilon_{t-j-1}\epsilon_{t-i-1})\right| < +\infty$

Proposition: If $(\epsilon_t)_{t\in\mathbb{Z}}$ is stationary ergodic and sMd process wrt some filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ and $\sum_{j=0}^{\infty} b^{2j} E(\epsilon_t^2 \epsilon_{t-j-1}^2) < +\infty$ and $\left| \sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} b^j b^i E(\epsilon_t^2 \epsilon_{t-j-1}\epsilon_{t-i-1})\right| < +\infty$ then $(\epsilon_t x_{t-1})_{t\in\mathbb{Z}}$ is a stationary ergodic and sMd process wrt some filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ with $E(\epsilon_0^2 x_{-1}^2) = \sum_{j=0}^{\infty} b^{2j} E(\epsilon_0^2 \epsilon_{-j-1}^2) + \sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} b^i b^j E(\epsilon_0^2 \epsilon_{-j-1}\epsilon_{-i-1})$ and due to the sMd CLT

$$\frac{1}{\sqrt{T}} \sum_{t=0}^{T} \epsilon_t x_{t-1} \xrightarrow{d} z \sim N\left(0, E(\epsilon_0^2 x_{-1}^2)\right) \text{ as } T\to\infty.$$

So remember that the deviation of the estimator from the true parameter, appropriately scaled is:

$$\sqrt{T}(b_T - b) = \frac{1}{1/T \sum_{t=1}^{T} x_{t-1}^2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t x_{t-1}$$

By Birkoff's LLN : $\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 \rightarrow E(x_{-1}^2) = \frac{E(\epsilon_0^2)}{1-\theta^2}$ a.s.

$\Rightarrow$ converges also in distribution to the same limit (degenerate r.v - a real number)

and we showed

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \epsilon_t x_{t-1} \xrightarrow{d} z$$

Therefore, jointly

$$\begin{pmatrix} 1/\sqrt{T} \sum_{t=1}^{T} \epsilon_t x_{t-1} \\ 1/T \sum_{t=1}^{T} x_{t-1}^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} z \\ \frac{E(\epsilon_0^2)}{1-\theta^2} \end{pmatrix}$$

and since the denominator is strictly positive, by CMT :

$$\frac{1}{1/T\sum_{t=1}^{T} x_{t-1}^2}\, \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \epsilon_t x_{t-1} \xrightarrow{d} \frac{z}{\frac{E(\epsilon_0^2)}{1-\theta^2}} = \frac{1-\theta^2}{E(\epsilon_0^2)}\, z$$

Hence,

$$\sqrt{T}(\theta_T - \theta) \xrightarrow{d} \frac{1-\theta^2}{E(\epsilon_0^2)}\, z \sim N\left(\cancel{\frac{1-\theta^2}{E(\epsilon_0^2)}} 0,\ \left[\frac{1-\theta^2}{E(\epsilon_0^2)}\right]^2 E(\epsilon_0^2 x_{-1}^2)\right)$$

$\hookrightarrow$ The rate of the estimator - the speed that θ_T converges to θ is $\sqrt{T}$.

<u>Theorem.</u> If $(\epsilon_t)_{t\in\mathbb{Z}}$ is stationary ergodic and an SMd process wrt some filtration, $E(\epsilon_0^2 x_{-1}^2) < +\infty$ and $|\theta| < 1$, then in the context of the AR(1) process constructed by $(\epsilon_t)_{t\in\mathbb{Z}}$ and θ, the following limit theory about the OLSE for θ holds

$$\sqrt{T}(\theta_T - \theta) \xrightarrow{d} N\left(0,\ \frac{(1-\theta^2)^2}{\left(E(\epsilon_0^2)\right)^2} E(\epsilon_0^2 x_{-1}^2)\right)$$

Does this hold when $\epsilon_t = \prod_{j=0}^{R} z_{t-j}$, as in the example that we have examined earlier?

We have already proved that it is a stationary ergodic SMd process, but we don't know anything about the existence of $E(\epsilon_0^2 x_{-1}^2)$, we need to specify it.

Assume for simplicity that $E(z_0^2) = 1$.

1. $R = 0$ hence $(\epsilon_t)_{t\in\mathbb{Z}}$ is iid with zero mean and finite variance (equal to one).

$$E(\epsilon_0^2 x_{-1}^2) = \sum_{j=0}^{\infty} \beta^{2j} E(\epsilon_0^2 \epsilon_{-1-j}^2) + \sum_{\substack{i,j=0\\ i\neq j}}^{\infty} \beta^i \beta^j E(\epsilon_0^2 \epsilon_{-1-i}\epsilon_{-1-j})$$

$$\overset{\text{indep.}}{=} \sum_{j=0}^{\infty} \beta^{2j} \underset{1}{\underset{\shortparallel}{E(\epsilon_0^2)}} \underset{1}{\underset{\shortparallel}{E(\epsilon_{-1-j}^2)}} + \underbrace{\sum_{\substack{i,j=0\\ i\neq j}}^{\infty} \beta^i\beta^j \underset{1}{\underset{\shortparallel}{E(\epsilon_0^2)}} \underset{0}{\underset{\shortparallel}{E(\epsilon_{-1-i})}} \underset{0}{\underset{\shortparallel}{E(\epsilon_{-1-j})}}}_{\text{exists and is equal to } 0}$$

$$= \sum_{j=0}^{\infty} \beta^{2j} = \sum_{j=0}^{\infty} (\beta^2)^j \overset{|\beta|<1}{=} \frac{1}{1-\beta^2}$$

Thereby, the variance of the asymptotic distribution of $\sqrt{T}(\beta_T - \beta)$ is $(1-\beta^2)^2 \dfrac{1}{1-\beta^2} = 1-\beta^2$

Hence, $\sqrt{T}(\beta_T - \beta) \xrightarrow{d} N(0, 1-\beta^2)$

2. $R > 0$

Let's look at each one of the series.

- $\sum_{\substack{i,k=0\\ i\neq k}}^{\infty} \beta^i \beta^k E(\epsilon_0^2 \epsilon_{-1-i}\epsilon_{-1-k})$

$$E(\epsilon_0^2 \epsilon_{-1-i}\epsilon_{-1-k}) = E\left[\left(\prod_{j=0}^{R} z_{-j}\right)^2 \prod_{j=0}^{R} z_{-1-i-j} \prod_{j=0}^{R} z_{-1-k-j}\right] \qquad (*)$$

some common elements exist $\Rightarrow$ a fourth moment of z will appear, so we must assume that it exists.

$E(z_0^4) = k \geq 1$ (since $E(z_0^2) = 1$ by Jensen's inequality)

Suppose without loss of generallity that $k > i$, so the latest z in the previous expectation is z_{-1-k-R} and we can write it as:

$$(*) \overset{\text{indep.}}{=} E\left(\prod_{j=0}^{R} z_{t-j}^2 \prod_{j=0}^{R} z_{-1-i-j} \prod_{j=0}^{R-1} z_{-1-k-j}\right) \cdot \underset{0}{\underset{\shortparallel}{E(z_{-1-k-R})}} = 0 \quad \forall\, i,k \ (i\neq k)$$

So when 4th moments exist:

$$\sum_{\substack{i,k=0\\ i\neq k}}^{\infty} \beta^i \beta^k \epsilon_{-1-i}\,\epsilon_{-1-k} = 0$$

Now, the first part

$$\cdot \sum_{h=0}^{\infty} \beta^{2h} E(\epsilon_0^2 \epsilon_{-1-h}^2) \quad (**)$$

$$E(\epsilon_0^2 \epsilon_{1-h}^2) = \underbrace{E\left(\prod_{j=0}^{R} z_{-j}^2 \prod_{j=0}^{R} z_{1-h-j}^2\right)}_{\text{4th elements of } z \text{ appear}} = \kappa^{\max(0,\,R-h)}$$

↳ Complete the proof

- what is common in the exp ⇒ Exp = 1
- What is not common ⇒ Exp = κ

Hint: try to find $E(\epsilon_0^2 / \mathcal{F}_{-1-h})$ for any h ⇓ try to apply LIE to (**)

$$\text{then } \sum_{h=0}^{\infty} \beta^{2h} (\epsilon_0^2 \epsilon_{t-h}^2) = \sum_{h=0}^{\infty} \beta^{2h} \kappa^{\max(0,R-h)} = \sum_{h=0}^{R-1} \beta^{2h} \kappa^{R-h} + \sum_{h=R}^{\infty} \beta^{2h} \cdot \kappa^0$$

$$= \kappa^R \sum_{h=0}^{R-1} \beta^{2h} \kappa^{-h} + \sum_{h=R}^{\infty} \beta^{2h} = \kappa^R \sum_{h=0}^{R-1} \left(\frac{\beta^2}{\kappa}\right)^h + \sum_{h=0}^{\infty} \beta^{2h} - \sum_{h=0}^{R-1} \beta^{2h} \quad (***)$$

Remember that, $\sum_{h=0}^{R-1} a^h = \frac{1-a^R}{1-a}$ if $a \neq 1$.

Since $\frac{\beta^2}{\kappa} \neq 1$ $(|\beta| < 1,\ \kappa \geq 1)$

$$(***) = \kappa^R \frac{1-(\beta^2/\kappa)^R}{1-\beta^2/\kappa} + \frac{1}{1-\beta^2} - \frac{1-(\beta^2)^R}{1-\beta^2} = \frac{\kappa^R - \frac{\beta^{2R}}{\kappa^R}\kappa^R}{\frac{1}{\kappa}(\kappa-\beta^2)} + \frac{(\beta^2)^R}{1-\beta^2}$$

$$= \frac{\kappa(\kappa^R - \beta^{2R})}{\kappa-\beta^2} + \frac{\beta^{2R}}{1-\beta^2}$$

$$\text{Thus, } E(\epsilon_0^2 x_{-1}^2) = \frac{\kappa(\kappa^R - \beta^{2R})}{\kappa-\beta^2} + \frac{\beta^{2R}}{1-\beta^2}$$

$$\text{and thereby, } \sqrt{T}(\beta_T - \beta) \xrightarrow{d} \mathcal{N}\left(0,\ (1-\beta^2)^2 \frac{\kappa(\kappa^R - \beta^{2R})}{\kappa-\beta^2} + (1-\beta^2)\beta^{2R}\right)$$

When R=0 we come back to the previous expression.

↳ Asymptotic properties might change when conditional heterosk might change.

↳ Statistical inference may have poor properties under conditional heteroskedasticity.

29/3/16

ARMA models

Why should we care about ARMA models or more in general for linear processes?

▶ Wold decomposition Theorem. If $y=(y_t)_{t\in\mathbb{Z}}$ is weakly stationary, then y can be decomposed as a finite pointwise sum of two sequences

$y_t = z + l$; $z=(z_t)_{t\in\mathbb{Z}}$ is deterministic (i.e a double real sequence)

and $l=(l_t)_{t\in\mathbb{Z}}$ is a linear process wrt a white noise process and a sequence of square summable coefficients.

(Any weakly stationary process has a deterministic and a stochastic component - linear process - that's why we examine them)

Remember, y is a linear causal process wrt $\epsilon=(\epsilon_t)_{t\in\mathbb{Z}} \sim WN(\sigma^2)$, $(b_j)_{j\in\mathbb{N}}$ is absolutely summable iff y_t can be expressed as

$$y_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} = \sum_{j=0}^{\infty} b_j L^j \epsilon_t = \varphi(L)\epsilon_t \quad \text{where} \quad \varphi(L) = \sum_{j=0}^{\infty} b_j L^j$$

Definition. Let $(b_j)_{j\in\mathbb{N}}$ be a real sequence then $\varphi(L) = \sum_{j=0}^{\infty} b_j L^j$ is called a formal power series wrt L or a linear filter.

Definition. If for some $p\in\mathbb{N}$, $b_j = 0 \ \forall\ j>p$ then $\varphi(L)$ is of finite order (it is essentially a polynomial wrt L). The minimal p for which this holds is called order of φ.

Example. $1 + b_1 L$, $\overset{0}{\neq}$ is this a formal power series wrt L?

Yes, it is since $b_j = \begin{cases} 1, & j=0 \\ b_1, & j=1 \\ 0, & j>1. \end{cases}$

and is of order 1.

Example. $1 - b_2 L^2$, $b_2 \neq 0$, $b_j = \begin{cases} 1, & j=0 \\ 0, & j=1 \\ b_2, & j=2 \\ 0, & j>2 \end{cases}$ So it is a formal power series wrt L.

Example. $b \neq 0$, $\sum_{j=0}^{\infty} b^j L^j$ it is a formal power series but not finite.

Definition. Consider that we have two formal power series $\varphi(L) = \sum_{j=0}^{\infty} b_j L^j$, $\psi(L) = \sum_{j=0}^{\infty} c_j L^j$, then we can define their product $\varphi(L) \cdot \psi(L)$ which is again a formal power series wrt L, say $\sum_{j=0}^{\infty} \gamma_j L^j$ where

$\gamma_j = \sum_{k=0}^{j} b_k c_{j-k}$ ← convolution of the 2 formal power series

$\gamma_0 = b_0 c_0$

$\gamma_1 = b_0 c_1 + c_0 b_1$

$\gamma_2 = b_0 c_2 + b_1 c_1 + b_2 c_0$

$\vdots$

Lemma 1. $\varphi(L)\psi(L) = \psi(L)\varphi(L)$ ↳ Exersice: proof Lemma 1

Lemma 2. Suppose that $\exists A > 0$, $b \in (0,1)$ for which $|b_j| \leq A b^j \ \forall j \in \mathbb{N}$ (which implies that they are absolutely summable $\sum_{j=0}^{\infty} |b_j| \leq \sum_{j=0}^{\infty} A b^j = A \sum_{j=0}^{\infty} b^j = \frac{A}{1-b}$) and ψ is of finite order.
Then $\exists A^* > 0$, $b^* \in (0,1)$ for which $|\gamma_j| \leq A^* b^{*j} \ \forall j \in \mathbb{N}$ (which guarantes absolute summability of the product)

↳ Exersice: proof Lemma 2

Definition. [Multiplicative inverse] $\psi(L)$ is a multiplicative inverse of φ iff $\psi(L)\varphi(L) = 1$ (denote it by φ^{-1})

FACT 1. [Uniqueness when exist] If ψ exists then it is unique.
Proof: Suppose that ψ^* is also an inverse of φ which implies that

$$\left.\begin{array}{l} \psi^*(L)\varphi(L) = 1 \text{ which implies that } \psi^*(L)\varphi(L)\psi(L) = \psi(L). \\ \text{But also } \varphi(L)\psi(L) = 1 \Rightarrow \psi^*(L)\varphi(L)\psi(L) = \psi^*(L) \end{array}\right\} \psi(L) = \psi^*(L)$$

Lemma 3. φ^{-1} exists iff $b_0 \neq 0$. Then (if $\varphi^{-1}(L) = \sum_{j=0}^{\infty} \varphi_j L^j$) using the fact that $\underbrace{\varphi(L)\varphi^{-1}(L)}_{\text{linear filters}} = 1$ $\rightarrow$ equality of linear filters $\Rightarrow$ equality of coefficients $\Rightarrow$ system of infinite equations on infinite unknowns.

$\varphi_0 = \frac{1}{b_0}$

$\varphi_j = -\frac{1}{b_0} \sum_{p=1}^{j} b_j \varphi_{j-p}$ (recurrence formula)

Lemma 4. Suppose that φ is of order p (i.e. it is of finite order and has the form $b_0 + b_1 L + b_2 L^2 + \ldots + b_p L^p$)

Consider the polynomial $b_0 + b_1 z + b_2 z^2 + \ldots + b_p z^p$, z is a complex variable. If the p roots of the previous polynomial have modulus greater than 1 (i.e. $|z^*| > 1$ where z^* is any such root or equivallently all the roots lie outside the unit disc of the complex plane) then there exist $A > 0$, $b \in (0,1)$ so that $|\varphi_j| \leq A b^j \ \forall j \in \mathbb{N}$

coefficients of the inverse

Therefore, the inverse has absolutely summable coefficients.

Example. $b_1 \neq 0$, $\varphi(L) = 1 - b_1 L$ which is a formal power series.

Does it have an inverse? By Lemma 3. exists a formal inverse.

The characteristic polynomial is $1 - b_1 z$, with root $1 - b_1 z = 0 \Rightarrow z = \frac{1}{b_1}$.

By Lemma 4. let's examine $\left|\frac{1}{b_1}\right|$. If $\left|\frac{1}{b_1}\right| > 1$ then the inverse has absolutely summable coefficients and have a certain form.

$\left|\frac{1}{b_1}\right| > 1 \Leftrightarrow |b_1| < 1$. !

Suppose then that $|b_1| < 1$, what is the inverse of φ? $\varphi^{-1}(L) = (1 - b_1 L)^{-1}$?

$\varphi_j = b_1^j \ \forall j \in \mathbb{N}$, $\varphi^{-1}(L) = (1 - b_1 L)^{-1} = \sum_{j=0}^{\infty} b^j L^j$

Example. $b_2 \neq 0$, $\varphi(L) = 1 - b_2 L^2$ which is a linear filter wrt L = formal power series.

↳ It has an inverse since $b_0 \neq 0$.

$\hookrightarrow 1 - b_2 z^2 = 0 \Leftrightarrow z^2 = \dfrac{1}{b_2}$ if $b_2 > 0 \to 2$ real roots $\pm\sqrt{\dfrac{1}{b_2}}$

$b_2 < 0 \to 2$ imaginary roots $\pm\sqrt{\dfrac{1}{|b_2|}}\, i$

. $\left|\pm\sqrt{\dfrac{1}{b_2}}\right| > 1 \Leftrightarrow \sqrt{\dfrac{1}{b_2}} > 1 \Leftrightarrow |b_2| < 1$, $b_2 \in (0,1)$

. $\left|\pm\sqrt{\dfrac{1}{|b_2|}}\, i\right| > 1 \Leftrightarrow \sqrt{\dfrac{1}{|b_2|}} > 1 \Leftrightarrow |b_2| < 1$, $b_2 \in (-1,0)$

then $\varphi_0 = 1$

$$\varphi_1 = -1 \sum_{p=1}^{1} b_1 \varphi_0 = 0$$

$$\varphi_2 = -1 \sum_{p=1}^{2} b_p \varphi_{2-p} = -(b_1 \varphi_1 + b_2 \varphi_0) = -(-b_2) = b_2$$

$\varphi_3 = 0$

$\varphi_4 = b_2^2$

$\varphi_5 = 0$

$\varphi_6 = b_2^3$

So, $\varphi_j = \begin{cases} 0, & j \text{ odd} \\ b_2^{\,j/2}, & j \text{ even} \end{cases}$

$$\varphi^{-1}(L) = \sum_{j=0}^{\infty} \varphi_j L^j = \sum_{j=0,2,4} b_2^{\,j/2} L^j \overset{\substack{k=j/2 \Leftrightarrow \\ j=2k}}{=\!=} \sum_{k=0}^{\infty} b_2^{\,k} L^{2k}$$

inverse of infinite order (while the series we started was of finite order)

▸ Let $\epsilon = (\epsilon_t)_{t \in \mathbb{Z}}$ is a white noise process and 2 filters of finite order $\varphi(L)$ and $\theta(L)$ which are of order p and q respectively.
Suppose that $\varphi(L) = 1 + \sum_{j=1}^{\infty} b_j L^j$ and $\theta(L) = 1 + \sum_{j=1}^{\infty} \theta_j L^j$

<u>Definition</u>. An ARMA process of order (p,q) [ARMA (p,q)] is any solution to the following stochastic recurrence equation $\varphi(L) y_t = \theta(L) \epsilon_t \Leftrightarrow$

$$\left(1 + \sum_{j=1}^{p} b_j L^j\right) y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \epsilon_t \Leftrightarrow y_t + \sum_{j=1}^{p} b_j y_{t-j} = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

So if $\varphi(L)$ has an inverse which is a linear filter and if we ensure that the product $\varphi^{-1}(L)\,\theta(L)$ is also a linear filter which satisfies some bounds which implies absolute summability $\Rightarrow$ the aforementioned recurrence equation has a solution (using Lemma 4 and 2), constructed upon the new linear filter $\varphi^{-1}(L)\,\theta(L)$

So,

<u>Theorem</u> If the roots of φ lie outside the unit disc of complex plane then there exists a solution of the form $y=(y_t)_{t\in\mathbb{Z}}$ with

$$y_t = \varphi^{-1}(L)\,\theta(L)\,\epsilon_t = \sum_{j=0}^{\infty}\delta_j L^j \epsilon_t = \sum_{j=0}^{\infty}\delta_j \epsilon_{t-j} \quad \forall\, t\in\mathbb{Z}$$

(δ_j satisfies the boundary conditions due to lemmas 4 and 2)

and there exist $A>0$, $b\in(0,1)$: $|\delta_j| < Ab^j$, $\forall\, j\in\mathbb{N}$ hence, $(\delta_j)_{j\in\mathbb{N}}$ is an absolutely summable sequence.

And thereby, y is weakly stationary and short memory (which implies that it is also regular by construction)

If moreover, $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic then $(y_t)_{t\in\mathbb{Z}}$ is also strictly stationary and ergodic

↳ So if all the roots of the AR part lie outside the unit disc then the aforementioned recurrence equation has a solution, which is a process consisting of absolutely summable coefficients, and is weakly stationary and short memory.

<u>Sketch of proof:</u> If the condition of the roots of φ holds $\overset{L4}{\Rightarrow}$ the φ^{-1} has coefficients that obey conditions of the form $|\varphi_j| \leq A^* b^{*j}$ $(A^*>0,\ b^*\in(0,1))$ $\overset{L2}{\Rightarrow}$ $\varphi^{-1}(L)\,\theta(L)$ is also a linear filter with coefficients that obey analogous conditions i.e $|\delta_j| \leq Ab^j$ $(A>0,\ b\in(0,1))$. →

Obviously, $y_t = \varphi^{-1}(L)\theta(L)\epsilon_t \quad \forall\, t \in \mathbb{Z}$, is a solution of the recurrence and due to the previous this is a linear causal process with coefficients that are absolutely summable.

And we have already established from the properties of such processes that y is weakly stationary with $E(y_t)=0 \quad \forall\, t \in \mathbb{Z}$ and $\gamma_K = \sum_{j=0}^{\infty} \delta_j \delta_{j+K}$

$$\sum_{K=0}^{\infty} |\gamma_K| = \sum_{K=0}^{\infty} \left| \sum_{j=0}^{\infty} \delta_j \delta_{j+K} \right| = \sum_{K=0}^{\infty}\sum_{j=0}^{\infty} |\delta_j \delta_{j+K}| = \sum_{K=0}^{\infty}\sum_{j=0}^{\infty} |\delta_j||\delta_{j+K}|$$

$$\leq \sum_{K=0}^{\infty}\sum_{j=0}^{\infty} Ab^j\, Ab^{j+K} = A^2 \sum_{K=0}^{\infty}\sum_{j=0}^{\infty} b^j b^{j+K} = A^2 \underbrace{\sum_{K=0}^{\infty} b^K}\ \underbrace{\sum_{j=0}^{\infty} b^{2j}}$$

2 independent series

$$= A^2 \frac{1}{1-b}\,\frac{1}{1-b^2} < +\infty$$

Therefore, the process is short memory.

(Short memory is guaranted by the boundary condition of the filters)

The final assertation follows from previous results on linear causal processes with absolutely summable coefficients.

- When $p=0 \Rightarrow \varphi$ has no roots $\Rightarrow$ the theorem holds trivially.

~~$\Rightarrow q=0 \rightarrow ARMA(1,0) = AR(1)$~~

Example $\theta(L)=1$, $\varphi(L) = 1-b_1L$

$(1-b_1L)y_t = \epsilon_t \Leftrightarrow y_t - b_1 y_{t-1} = \epsilon_t$

Then $y_t = \varphi^{-1}(L)\epsilon_t = \sum_{j=0}^{\infty} b_1^j L^j \epsilon_t = \sum_{j=0}^{\infty} b_1^j \epsilon_{t-j}$

Exercise: $\theta(L)=1$, $\varphi(L) = 1-b_2L^2$
- form of the solution
- autocovariance fnc

Exercise: $\theta(L) = 1+\theta_1 L$, $\theta_1 \neq 0$, $\varphi(L) = 1-b_1L$
- solution
- autocovariance fnc

5/4/16

Example. Consider an ARMA(1,1), $\varphi(L) = 1 - \phi_1 L$, $\theta(L) = 1 - \theta_1 L$

We are seeking the ARMA(1,1) recurrence and the respective solution with the analogous properties.

$(1-\phi_1 L) y_t = (1-\theta_1 L)\epsilon_t \quad \forall t \in \mathbb{Z}$

$\Leftrightarrow y_t - \phi_1 y_{t-1} = \epsilon_t - \theta_1 \epsilon_{t-1} \quad \forall t \in \mathbb{Z}$

$\Leftrightarrow y_t = \phi y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \quad \forall t \in \mathbb{Z}$

Remember that we need $|\phi| < 1$, but we haven't established any theory which includes restrictions for the roots of the θ polynomial.

So unit disc condition (UDC) $\Leftrightarrow |\phi| < 1$.

If UDC holds, then we can define y_t as:

$$y_t = \sum_{j=0}^{\infty} \phi_1^j L^j (1-\theta_1 L)\epsilon_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} - \theta_1 \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j-1}$$

$$= \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} - \theta_1 \sum_{j=0}^{\infty} \phi_1^{j-1} \epsilon_{t-j} = \epsilon_t + \sum_{j=1}^{\infty} (\phi_1^j - \theta_1 \phi_1^{j-1})\epsilon_{t-j}$$

$$= \epsilon_t + (\phi_1 - \theta_1) \sum_{j=0}^{\infty} \phi_1^{j-1} \epsilon_{t-j} = \sum_{j=0}^{\infty} \phi_j^* \epsilon_{t-j} \quad \text{where } \phi_j^* = \begin{cases} 1, & j=0 \\ (\phi_1-\theta_1)\phi_1^{j-1}, & j>0 \end{cases}$$

- If $\phi_1 = \theta_1$ (i.e. if the polynomials share a common root) then the process is essentially a WN process.

(If $\varphi(q)$ and $\theta(p)$ share a common root $\Rightarrow \varphi(q-1)$, $\theta(p-1)$

$\Rightarrow$ identification problem $\rightarrow$ Box-Jenkins methodology

$\Rightarrow$ we select as our statistical model the ARMA of the lower order)

Since y_t is a linear process we know that:

$$\gamma_0 = E(\epsilon_0^2) \sum_{j=0}^{\infty} \phi_j^{*2} = E(\epsilon_0^2)\left(\phi_0^{*2} + \sum_{j=1}^{\infty} \phi_j^{*2}\right) = E(\epsilon_0^2)\left[1 + \sum_{j=1}^{\infty} \left((\phi_1-\theta_1)\phi_1^{j-1}\right)^2\right]$$

$$= E(\epsilon_0^2)\left[1 + (\phi_1-\theta_1)^2 \sum_{j=1}^{\infty} \phi_1^{2j-2}\right] = E(\epsilon_0^2)\left[1 + \frac{(\phi_1-\theta_1)^2}{\phi_1^2} \sum_{j=1}^{\infty} (\phi_1^2)^j\right]$$

$$= E(\epsilon_0^2)\left(1 + \frac{(\phi_1-\theta_1)^2}{\cancel{\phi_1^2}} \frac{\cancel{\phi_1^2}}{1-\phi_1^2}\right) = E(\epsilon_0^2)\left(1 + \frac{(\phi_1-\theta_1)^2}{1-\phi_1^2}\right)$$

and $$\gamma_k = E(\epsilon_0^2) \sum_{j=0}^{\infty} \phi_j^* \phi_{j+k}^* = E(\epsilon_0^2)\left[(\phi_1-\theta_1)\phi_1^{k-1} + \sum_{j=1}^{\infty} (\phi_1-\theta_1)\phi_1^{j-1}(\phi_1-\theta_1)\phi_1^{j+k-1}\right]$$

$$= E(\epsilon_0^2)\left[(\phi_1-\theta_1)\phi_1^{k-1} + (\phi_1-\theta_1)^2 \sum_{j=1}^{\infty} \phi_1^{2j-2+k}\right]$$

$$= E(\epsilon_0^2)\left[(\phi_1-\theta_1)\phi_1^{k-1} + (\phi_1-\theta_1)^2 \phi_1^{k-2} \sum_{j=1}^{\infty} (\phi^2)^j\right]$$

$$= E(\epsilon_0^2)\left[(\phi_1-\theta_1)\phi_1^{k-1} + (\phi_1-\theta_1)^2 \phi_1^{k-2} \frac{\phi_1^2}{1-\phi_1^2}\right]$$

$$= E(\epsilon_0^2)\left[(\phi_1-\theta_1)\phi_1^{k-1} + (\phi_1-\theta_1)^2 \frac{\phi_1^k}{1-\phi_1^2}\right]$$

generalizes the results for AR(1), MA(1), WN processes.

$\theta_1 = 0 \Rightarrow$ AR(1)

$\phi_1 = 0 \Rightarrow$ MA(1)

$\theta_1 = 0,\ \phi_1 = 0 \Rightarrow$ WN

↳ Exercise: ARMA(1,2). Find γ_0, γ_k.

Invertibility

Definition: The ARMA (p,q) process is called invertible iff there exists a sequence $(\psi_j)_{j\in\mathbb{N}}$: $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that $\epsilon_t = \sum_{j=0}^{\infty} \psi_j y_{t-j}$ $\forall\ t \in \mathbb{Z}$

(i.e. such that the WN process can be expressed as a linear process wrt the ARMA process)

Remark. Invertibility directly implies that $(\epsilon_t)_{t\in\mathbb{Z}}$ is adopted to $(f_t)_{t\in\mathbb{Z}}$ where $f_i := \sigma(y_{t-i},\ i \geq 0)$.

Example. AR(p), $y_t \underbrace{\left(1 - \sum_{j=1}^{p} \phi_j L^j\right)}_{\varphi(L)} = \epsilon_t \Leftrightarrow y_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j} = \epsilon_t$ $\forall\ t \in \mathbb{Z}$

So is it invertible? Can we write it as $\sum_{j=0}^{\infty} \psi_j y_{t-j} = \epsilon_t$ $\forall\ t \in \mathbb{Z}$?

Yes, for $\psi_0 = 1$, $\psi_j = -\phi_j$, $j = 1, \dots, p$, $\psi_j = 0$, $j > p$

for which $\sum_{j=0}^{\infty} |\psi_j| < \infty$ since it is essentially a finite sum (since for $j > p$, $\psi_j = 0$)

Therefore, by construction every $AR(p)$ is invertible.

- $y_t \varphi(L) = \theta(L)\epsilon_t \ \forall t \in \mathbb{Z}$. So when this is invertible?

  When the roots of θ satisfy the unit disc condition, the process is invertible.

<u>Theorem.</u> If the roots of θ satisfy the UDC then the ARMA (p,q) process is also invertible.

<u>Example.</u> MA(1). $y_t = (1-\theta_1 L)\epsilon_t \ \forall t \in \mathbb{Z}$

$\Leftrightarrow y_t = \epsilon_t - \theta_1 \epsilon_{t-1} \ \forall t \in \mathbb{Z}$ Obviously, this is a linear process on the ϵ, with absolutely summable coefficients.

If $|\theta_1| < 1$ the process is invertible and $\theta_1^{-1}(L) = \sum_{j=0}^{\infty} \theta_1^j L^j$

Thereby, $\epsilon_t = \theta(L)^{-1} y_t = \sum_{j=0}^{\infty} \theta_1^j L^j y_t$

$= \sum_{j=0}^{\infty} \theta_1^j y_{t-j}$ so in this case $\psi_j = \theta_1^j \ \forall j = 0, 1, \dots$

<u>Exercise:</u> Find conditions for invertibility for $MA(2)$. Find the ψ_j's

<u>Discussion on issues of Estimation and Inference for ARMA processes.</u>

If we have MA component, the OLS is not feasible since regressors contain parts that are not observable.

Assume that, it is available a sample (a realization of the random variable) i.e. a random vector of size $T+p-1$, $(y_t)_{t=-p+1,\dots,1,\dots,T}$

Assume that we know p and q and we only need to find the parameters of the relevant polynomials $\varphi(L) = 1 - \sum_{j=1}^{p} \phi_{j_0} L^j$, $\theta(L) = 1 - \sum_{j=1}^{q} \theta_{j_0} L^j$.

for $\phi_{1_0}, \phi_{2_0}, \dots, \phi_{p_0}, \theta_{1_0}, \theta_{2_0}, \dots, \theta_{q_0}$ unknown

$\psi = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q) \in \Theta \subseteq \mathbb{R}^{p+q}$

Condition: $\varphi_0 \in \Theta$ in order to have well-specification.

- Case 1: $q=0$ ($AR(p)$ - case)

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}$$

Can we do OLS?

$Y = X\varphi' + \epsilon$ where $X = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-p+1} \\ y_1 & y_0 & \cdots & y_{-p+2} \\ \vdots & & & \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{bmatrix}_{T\times p}$

and $\varphi = (b_1, b_2, \ldots, b_p)$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

Rank condition: $T \geq p$.

Assume that $\Theta = \mathbb{R}^p$ then $\varphi_T' = (X'X)^{-1}X'Y$

Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ is also strictly stationary and ergodic.

A direct implication of the fact that the particular statistical model is well-specified is that:

$$\varphi_T' = \varphi_0' + (X'X)^{-1}X'\epsilon = \varphi_0' + \left(\frac{X'X}{T}\right)^{-1}\frac{X'\epsilon}{T}$$

- $\frac{X'\epsilon}{T} = \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t y_{t-i}\right)_{i=1,\ldots,p}}_{i^{th} \text{ element of the vector } \frac{X'\epsilon}{T}}$ each element of this vector is a part of a stationary ergodic process.

So if $E|\epsilon_t y_{t-i}| < +\infty$ by Birkoff's LLN each element will converge to:

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t y_{t-i} \to E(\epsilon_t y_{t-i}) = 0 \text{ a.s.}$$

So jointly the vector: $\frac{X'\epsilon}{T} \to 0_{p\times 1}$ a.s.

- $\frac{X'X}{T} = \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-i}\, y_{t-j}\right)_{i,j=1,\ldots,p}}_{i,j^{th} \text{ element of } \frac{X'X}{T} \text{ matrix}}$

Any component of this matrix is stationary ergodic,
and for $i=j$, if $E(y_t^2)<+\infty$ by Birkoff's LLN :

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow E(y_{t-1}^2) \quad \text{a.s.}$$

$i \neq j$, if $E|y_t y_{t-j}| < +\infty$ by Birkoff's LLN :

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-i}\, y_{t-j} \longrightarrow E(y_{t-i}\, y_{t-j}) = \gamma_{|i-j|} \quad \text{a.s.}$$

So, jointly the matrix : $\frac{X'X}{T} \longrightarrow \underbrace{(\gamma_{|i-j|})_{i,j=1,\dots,p}}_{=\text{autocovariance matrix}}$

We can prove that since $\gamma_0 > 0$ and $\gamma_K \to 0$ as $K \to \infty$, this matrix will always be invertible.
Thereby, by the CMT : $\left(\frac{X'X}{T}\right)^{-1} \longrightarrow \left(\gamma_{|i-j|}\right)^{-1}_{i,j=1,\dots,p}$

and also by CMT :
$\varphi_T' \longrightarrow \varphi_0' + (\quad)^{-1} \cdot 0$ a.s.
$\varphi_T' \longrightarrow \varphi_0'$ a.s.

12/4/16

Case 2 : $q > 0$

Could we use an analogous linear model and compute an analogous OLS estimator? Since $q > 0$, some of the collumns in the regressor matrix contain ϵ which are unobservable, therefore we can't derive the OLS. Thereby, regressors matrix in such a context, contain latent variables so we can't perform OLS.

Gaussian Quasi-Likelihood Function

Consider a sample $(y_t)_{t=-p+1,\dots,T}$

The process is essentially a solution of such a recurrence equation:

$$y_t = \sum_{j=1}^{p} b_{j_0} y_{t-j} - \sum_{j=1}^{q} \theta_{j_0} \epsilon_{t-j} + \epsilon_t \Leftrightarrow$$

$$\epsilon_t = y_t - \sum_{j=1}^{p} b_{j_0} y_{t-j} + \sum_{j=1}^{q} \theta_{j_0} \epsilon_{t-j}$$

(We consider that the variance of the WN process is known and equal to 1)

Suppose that we have an extra structure for ϵ, suppose that ϵ_t are iid with standard normal marginal distribution.

Then we could derive that the joint distribution of $(y_t)_{t=1,\dots,T}$ given $y_0, y_{-1}, \dots, y_{1-p}, \epsilon_0, \epsilon_1, \dots, \epsilon_{1-q}$ has a form like this:

$$\frac{1}{(\sqrt{2\pi})^T} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{p} b_j y_{t-j} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}\right)^2\right)$$ function of the sample and the parameters

the maximization of which is equivalent to:

$$-\frac{1}{2T}\sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{p} b_j y_{t-j} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}\right)^2$$

The above is termed as conditional maximum likelihood estimator.

However, still ϵ are unobservable, we have the same problem in estimation

(observe that we essentially have a sum of squares criterion)

Now, assume that normallity and iidness of the ϵ's are not correct then the previous joint density does not represent the real density, it is not correct and it is actually termed as the Gaussian Quasi Maximum likelihood. Every maximizer of such a function is called QMLE.

The computation problems due to unobservable ϵ still exist.
We need an approximation for the ϵs.
Remember that the vector of parameters is $\varphi = (b_1, b_2, \dots, b_p, \theta_1, \dots, \theta_q) \in \Theta \subseteq \mathbb{R}^{p+q}$
If we could define a likelihood fnc as function of the sample and the values of the parameters only and not as a function of the unobservable ϵ, then we could derive a feasible estimator.
We will try to approximate ϵ's i.e. to describe a filtering algorithm.

Filtering algorithm for the unobserved components.
Given a value of φ:
1. Set $e^*_{-q+1} = e^*_{-q+2} = \dots = e^*_0 = 0$, e^*: approximated ϵ

2. $e^*_t(\varphi) := y_t - \sum_{j=1}^{p} b_j y_{t-j} + \sum_{j=1}^{q} \theta_j e^*_{t-j} \quad \forall\ t = 1, \dots, T$

so this is computable since it is a function of the sample and φ.

Thereby, now we can define a feasible criterion.
$$Q^*_T(\varphi) = -\frac{1}{2T} \sum_{t=1}^{T} \left(y_t - \sum_{j=1}^{p} b_j y_{t-j} + \sum_{j=1}^{q} \theta_j e^*_{t-j} \right)^2 = -\frac{1}{2T} \sum_{t=1}^{T} e_t^*(\varphi)$$

Definition: In this framework the Gaussian QMLE φ_T is defined by
$\varphi_T \in \underset{\varphi \in \Theta}{\operatorname{argmax}}\, Q_T^*(\varphi)$. (which is an M-estimator

If we try to maximize the previous then we can't obtain an analytical solution. We need a numerical algorithm.

Numerical Optimization of $Q^*_T(\varphi)$ wrt $\varphi \in \Theta$.

1. Initialize φ
2. For φ employ the filtering algorithm to compute $(e^*_t(\varphi))_{t=1,\dots,T}$ and $Q^*_T(\varphi)$.
3. Change φ according to information obtained from $Q^*_T(\varphi)$ in step 2.
4. Go back to 2-3 until some numerical criteria are fullfield.
5. φ_T equals to the value of φ in the final call of step 2.

(There are possible optimization errors in every step of the algorithm. For example, if we choose an initial value far from the true optimizer, we will possibly have a large optimization error.)

How can we be sure that the argmax is not empty for some sample values? For the sample values that this argmax is empty with positive probability the estimator can not be defined.

If Q is continuous, Θ is compact
If Q is concave, Θ is a convex subset of $\mathbb{R}^{p+q}$ $\Rightarrow$ argmax of the approximate minimizer exists.

Also, argmax must be measurable. There are many things concerning the existence of φ_T.

Example. $p=0$, $q=1$ (MA(1) case)

$\varphi_0=\vartheta_{1_0}$, we assume invertibility therefore we know that $\vartheta_{1_0} \in (-1,1)$

So $\varphi=\vartheta_1 \in \Theta=(-1,1)$

The Gaussian quasi-likelihood is denoted by:

$$Q^*_T(\vartheta_1) = -\frac{1}{2T}\sum_{t=1}^{T} e^{*}_t(\vartheta_1) = -\frac{1}{2T}\sum_{t=1}^{T}\left(y_t-\vartheta_1 e^*_{t-1}(\vartheta_1)\right)^2$$

The $(e^*_t(\vartheta_1))_{t=0}^{T}$ are defined according to the filtering algorithm as

For a value of θ_1.

1. $e_0^*(\theta_1) = 0$
2. $e_t^*(\theta_1) = y_t + \theta_1 e_{t-1}^*(\theta_1) \quad \forall\ t = 1, \dots, T$

Then, the Gaussian QMLE $\theta_{1_T} \in \underset{\theta_1 \in (-1,1)}{\operatorname{argmax}}\ Q_T^*(\theta_1)$

Assumption. $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

Let's take a look at $e_t^*(\theta_1)$.

$$e_t^*(\theta_1) = \underbrace{y_t}_{\text{true model MA(1)}} + \theta_1 e_{t-1}^*(\theta_1)$$

$$= \epsilon_t - \theta_{1_0}\epsilon_{t-1} + \theta_1 e_{t-1}^*(\theta_1) \quad \forall\ t=1,\dots,T$$

$\Leftrightarrow e_t^*(\theta_1) - \theta_1 e_{t-1}^*(\theta_1) = \epsilon_t - \theta_{1_0}\epsilon_{t-1}$ which is an ARMA(1,1) recursion (if you put $e^* = y$). So the filtering defines an ARMA(1,1). What whould be the properties of this process, if this process exists? (If UDC holds - which holds since we have assumed $\theta \in (-1,1)$) $\Rightarrow$ unique stationary ergodic solution exists.

When $\theta_{1_0} = \theta_1$ we have a common root $\Rightarrow$ reduces to WN process.

(somehow we can see the approximation)

Since $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic $\Rightarrow e_t^*(\theta_1)$ is str. stationary + ergodic $\forall \theta_1$

By Birkoff's LLN:

$$\frac{1}{T}\sum_{t=1}^{T} e_t^{*2}(\theta_1) \longrightarrow \mathbb{E}(\epsilon_0^{*2}(\theta_1))$$

variance of the ARMA(1,1) process

(we can show that it converges not only pointwise but also uniformly (locally uniformly) - ULLN)

$\longrightarrow$

Given a uniform version of Birkoff's LLN we can establish that $Q_T^*(\theta_1)$ converges almost surely and locally uniformly wrt Θ to

$$\boxed{1 + \frac{(\theta_1-\theta_1)^2}{1-\theta_1^2}} = \underbrace{\left(1 + \frac{\theta_1 - \theta_{10}}{1-\theta_1^2}\right)}_{\text{which is minimized at } \theta_1 = \theta_{10}}\left(-\frac{1}{2}\right) \Rightarrow \text{maximized at } \theta_1 = \theta_{10}$$

$\Rightarrow$ strongly consistent.

19/4/16

Baysian Information Criterion

Suppose that p,q are unknown up to a point (we only know some upper bounds)

$p_0 \leq p_{max}$

$q_0 \leq q_{max}$

For every possible combination of p,q, $\forall p,q : p \leq p_{max}, q \leq q_{max}$

we compute the QMLE $\varphi_T(p,q)$ and the likelihood function evaluated at QMLE

$-\frac{2}{T} \sum e_t^{*2}(\varphi_T(p,q)) := -2\,ASSR(\varphi_T(p,q))$

$$bic_T(p,q) := \ln ASSR(\varphi_T(p,q)) + (p+q)\frac{\ln T}{T}$$

Thereby, an estimator for p and q is the minimizer of $bic_T(p,q)$.

$$(p_T, q_T) := \underset{\substack{0 \leq p \leq p_{max} \\ 0 \leq q \leq q_{max}}}{\operatorname{argmin}} bic_T(p,q)$$

e.g. If $q_0 = 0$ known and $(\epsilon_t)_{t\in\mathbb{Z}}$ iid, $E(\epsilon_0^4) < +\infty$, we can prove that $p_T \xrightarrow{P} p_0$

An example of an Indirect Inference Estimator in the context of a MA(1) model

Suppose that $(y_t)_{t=0}^{T}$ is a part of a MA(1) process with $|\theta_{1_0}| < 1$ (hence invertible) (but θ_{1_0} is unknown)

$\Theta = (-1,1)$

Can we construct a linear model from our sample?

$Y = bX + e$

such that $Y = (y_t)_{t=1}^{T}$, $X = (y_t)_{t=0}^{T-1}$, $b \in \mathbb{R}$

which we can estimate by OLS : $b_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$ in the framework of the AR(1) statistical model.

However, our model is misspecified. The only case that our statistical model could be well-specified is if $q=0$, which will occur when $\theta_{1_0}=0$.

<u>Assumption</u>: $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

$$b_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = \frac{1/T \sum_{t=1}^{T} y_t y_{t-1}}{1/T \sum_{t=1}^{T} y_{t-1}^2}$$

Since (ϵ_t) is stationary ergodic $\Rightarrow (y_t)$ is also stationary ergodic $\Rightarrow (y_{t-1})$ stat. erg.
and so does their product $(y_t y_{t-1})$: stat. ergodic.
and (y_{t-1}^2) is also stat. ergodic.
Thereby, by Birkoff's LLN:

$$\frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} \longrightarrow \underset{\theta_{1_0}}{\underset{\shortparallel}{E(y_0 y_{-1})}} \quad \text{a.s.}$$

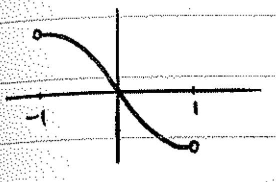
$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \underset{1+\theta_{1_0}^2 > 0}{\underset{\shortparallel}{E(y_{t-1}^2)}} \quad \text{a.s.}$$

Hence, $b_T \longrightarrow \dfrac{-\theta_{1_0}}{1+\theta_{1_0}^2}$ a.s. which is inconsistent, except from the case that $\theta_{1_0}=0$ which means that our auxiliary model (AR(1)) is well-specified (WN will fall within the case of AR(1))
However, this estimator gives us some information about θ_{1_0}.
Suppose that we knew the limits (e.g. 1/3) and also the form of the limit, could we find θ_{1_0}? Yes, solving $-\dfrac{\theta_{1_0}}{1+\theta_{1_0}^2} = \dfrac{1}{3} \Rightarrow \theta_{1_0}$

However, we don't know that limit, but we know something that approximates that limit, b_T.
So we could see if by equating b_T with an expression of $\dfrac{-\theta_{1_0}}{1+\theta_{1_0}^2}$, we can obtain an estimator of θ_{1_0} (analogy principle).
Thereby, $\dfrac{-\theta_{1_0}}{1+\theta_{1_0}^2} = \dfrac{-\theta_1}{1+\theta_1^2}$ by the analogy principle

we would solve $b_T = -\dfrac{\theta_1}{1+\theta_1^2}$

Let's call $b(\theta_1) = -\underbrace{\dfrac{\theta_1}{1+\theta_1^2}} : (-1,1) \to (-1/2, 1/2)$: binding function



limit of the OLSE of the MA(1), if the parameter was θ_1. So, essentially this function has arisen as a collection of all the limits of the OLSE in $(-1,1)$.

Instead of that, we could define some kind of distance between β_T and the binding function.

<u>Definition</u>: The Indirect Inference Estimator say θ_T is defined by

$$\theta_T \in \underset{\theta_1 \in \Theta}{\operatorname{argmax}} \left(\beta_T + \frac{\theta_1}{1+\theta_1^2} \right)^2 \quad \text{(using the Euclidean distance)}$$

Actually, we have an estimator in 2 steps.

Does this estimator exist? The aforementioned function is 1-1 [The fact that it is 1-1, implies that $b(\theta_1) = b(\theta_{1_0})$ has unique solution $\Leftrightarrow \theta_1 = \theta_{1_0}$: Indirect Identification]. So when $\beta_T \in (-1/2, 1/2)$ the optimization problem has unique solution.

$$b(\theta_1) = \beta_T \Leftrightarrow -\frac{\theta_1}{1+\theta_1^2} = \beta_T \Rightarrow -\theta_1 = \beta_T + \theta_1^2 \beta_T \Rightarrow \theta_1^2 \beta_T + \theta_1 + \beta_T = 0$$

$\Rightarrow$ 2 roots, the one of them will give $-\dfrac{\theta_1^2}{1+\theta_1^2} \in (-1/2, 1/2)$

$$\Rightarrow \theta_T = -1 + \frac{\sqrt{1-4\beta_T^2}}{2\beta_T}$$

We could define this estimator for every value of β_T, even outside the interval $(-1/2, 1/2)$. However, we don't really care since asymptotically $\beta_T \to b(\theta_{1_0}) \in (-1/2, 1/2)$ a.s. as $T \to \infty$, $\theta_{1,T}$ admits the previous expression.

$$\text{So, } \theta_T = -1 + \frac{\sqrt{1-4\beta_T^2}}{2\beta_T} = b^{-1}(\theta_T)$$

Consistency

Almost surely $\theta_T = b^{-1}(\delta_T)$ as $T \to \infty$

As we saw before $\delta_T \to b(\theta_{1_0})$ a.s.

By the CMT

since $b^{-1}: (-1/2, 1/2) \to \Theta$ is continuous

$$b^{-1}(\delta_T) \to b^{-1}(b(\theta_{1_0})) = \theta_{1_0} \text{ a.s.}$$

Hence, $\theta_T \to \theta_{1_0}$ a.s. as $T \to \infty$.

Assumption: $(\epsilon_t)_{t \in \mathbb{Z}}$ is iid.

We can find the asymptotic behavior of:

$$\sqrt{T}(\delta_T - b(\theta_{1_0})) = \sqrt{T}\left(\delta_T + \frac{\theta_{1_0}}{1+\theta_{1_0}^2}\right) \xrightarrow{d} Z \sim \mathcal{N}(0, V_{\theta_{1_0}})$$

$$\text{where } V_{\theta_{1_0}} = \frac{(1+\theta_{1_0}^4)^2 + \theta_{1_0}^2(1+\theta_{1_0}^2)^2}{(1+\theta_{1_0}^2)^4}$$

Therefore, can we find the asymptotic distribution of $\theta_T = b^{-1}(\delta_T)$?

By the delta method:

$$\sqrt{T}\left(F(\delta_T) - F(b_0)\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\partial F(b_0)}{\partial b}\right)^2 V\right)$$

So we are interested in

$$\sqrt{T}\left(b_1^{-}(\delta_T) - b_1^{-}(b(\theta_{1_0}))\right)$$

$=$ as $T \to \infty$

$$\sqrt{T}(\theta_T - \theta_0)$$

In order to use the delta method, b^{-1} must be continuously differentiable in an open neighborhood of $b(\theta_{1_0})$.

(We can either compute the derivative or use the inverse fnc th^{m}).

$$\frac{\partial b^{-1}(b(\theta_{1_0}))}{\partial b} = \frac{1}{\left.\frac{\partial b}{\partial \theta_1}\right|_{\theta_1 = \theta_{1_0}}}$$

inverse function th^{m}

$$\frac{\partial b(\theta_1)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1}\left(-\frac{\theta_1}{1+\theta_1^2}\right)\Bigg|_{\theta_1=\theta_{1_0}} = \frac{-(1+\theta_1^2)+\theta_1 2\theta_1^2}{(1+\theta_1^2)^2}\Bigg|_{\theta_1=\theta_{1_0}} = \frac{\theta_{1_0}^2-1}{(1+\theta_{1_0}^2)^2}$$

Hence, $\dfrac{\partial b^{-1}(b(\theta_{1_0}))}{\partial \theta} = \dfrac{(1+\theta_{1_0}^2)^2}{\theta_{1_0}^2-1}$

Therefore, $\sqrt{T}(\theta_T-\theta_{1_0}) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{(1+\theta_{1_0}^2)^4}{(\theta_{1_0}^2-1)^2} V_{\theta_{1_0}}\right)$

However, the Gaussian QMLE is more efficient. The advantage of the IE is the fact that it doesn't require any numerical procedure. Can we find a consistent estimator of the asymptotic variance? AVar is a continuous function of θ_{1_0} so we could use θ_T which is a consistent estimator of θ_{1_0}.

Conditionally Heteroskedastic Processes

Framework: $(z_t)_{t\in\mathbb{Z}}$ is iid with $E(z_0)=0$ and $E(z_0^2)=1$

Consider $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, $\mathcal{F}_t := \sigma(z_{t-1}, z_{t-2}, \dots) = \sigma(z_{t-1-i}, i \geq 0)$.

Suppose that $(h_t)_{t\in\mathbb{Z}}$ exists and is adopted to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ (h_t is a measurable function of the elements that constitute this σ-algebra) and $E(h_t) < +\infty$ $\forall t \in \mathbb{Z}$ (this assumption may not hold, but here we assume it holds and then h_t is a conditional variance process), $h_t \geq 0$ a.s. $\forall t \in \mathbb{Z}$.

Definition: The process $(y_t)_{t\in\mathbb{Z}}$, $y_t := z_t\sqrt{h_t}$ is called a conditionally (on $(\mathcal{F}_t)_{t\in\mathbb{Z}}$) heteroskedastic process, where h_t is the conditional on $\mathcal{F}_t$ variance of y_t $\forall t \in \mathbb{Z}$.

The fact that $E(h_t) < +\infty$ implies that also the square root has finite 1st moment, $E(\sqrt{h_t}) < +\infty$ $\forall t$

- $E(y_t / \mathcal{F}_t) = E(z_t\sqrt{h_t})$ Is it well-defined?

In order to be well-defined, the unconditional expectation must be well-defined.

$E(z_t\sqrt{h_t})$, h_t is adopted to f_t and z_t is iid

$= E(z_t)\underbrace{E(\sqrt{h_t})}_{<+\infty} = 0$

So $E(y_t) \overset{LIE}{=} E(E(y_t/f_t)) = 0$

- $V(y_t/f_t) = E(y_t^2/f_t)$ in order this to make sense, $E(y_t^2)$ must exist.

$E(y_t^2) = E(z_t^2h_t) \overset{z_t \text{ indep.}}{=} \cancel{E(z_t^2)}^{1}\underbrace{E(h_t)}_{<+\infty} = E(h_t) < +\infty \quad \forall\, t \in \mathbb{Z}$

So $V(y_t/f_t) = E(z_t^2h_t/f_t) = h_t\, E(z_t^2/f_t) = h_t \cdot E(z_t^2) = h_t$

then

$E(y_t^2) \overset{LIE}{=} E(E(y_t^2/f_t)) = E(h_t)$

10/5/16

Conditionally Heteroskedastic Processes

Assumption: If $(h_t)_{t\in \mathbb{Z}}$ is stationary and ergodic

What about (y_t) process now? Is it stationary and ergodic?

Yes, since (z_t) is an iid process thus strictly stationary and ergodic and $(\sqrt{h_t})$ is a continuous function of (h_t) so stationary ergodic. Thereby, their product is also stationary ergodic.

Is $(y_t)_{t\in \mathbb{Z}}$ a s.m.d. process?

- It is not adapted to $(f_t) = \sigma(z_{t-1-i}, i \geq 0)$ since $y_t = \underline{\underline{z_t}}\sqrt{h_t}$.

We can define another σ-algebra $(G_t)_{t\in \mathbb{Z}}$, $G := \sigma(z_{t-i}, i \geq 0)$ and observe that $f_t = G_{t-1}$

So $(y_t)_{t\in \mathbb{Z}}$ is adapted to $(G_t)_{t\in \mathbb{Z}}$.

- $E(y_t^2) < +\infty \quad \forall\, t \in \mathbb{Z}$
- $E(y_t / G_{t-1}) = E(y_t / f_t) = 0 \quad \forall\, t \in \mathbb{Z}$

So the previous framework (without the assumption) is enough to imply that (y_t) is a s.m.d..

All those things together imply that $(y_t)_{t\in \mathbb{Z}}$ is strictly stationary and ergodic s.m.d. process. So we created a process with those characteristics which at the same time is conditionally heteroskedastic (i.e. with a stochastic variance). The part of the framework that we haven't specified is the process $(h_t)_{t\in}$

GARCH (1,1)

Definition. Suppose that $\omega, a, b \in \mathbb{R}$. The $(y_t)_{t\in \mathbb{Z}}$ is termed a GARCH (1,1) process iff $(h_t)_{t\in \mathbb{Z}}$ satisfies the stochastic recursion:

$$h_t = \omega + a y_{t-1}^2 + b h_{t-1} = \omega + a z_{t-1}^2 h_{t-1} + b h_{t-1}$$

$$= \omega + (a z_{t-1}^2 + b) h_{t-1} \quad \forall\, t \in \mathbb{Z}$$

Is it possible solutions of the previous recursions to exist?

Suppose that $(h_t)_{t\in\mathbb{Z}}$ exists as a solution, then does positivity hold?

Positivity.

The previous recursion can be written as a quadratic form:

$$h_t = \begin{pmatrix} 1 & y_{t-1} & h_{t-1}^{1/2} \end{pmatrix} \underbrace{\begin{pmatrix} \omega & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{pmatrix}} \begin{pmatrix} 1 \\ y_{t-1} \\ h_{t-1}^{1/2} \end{pmatrix}$$

positivity is ensured if this matrix is P.D. and since it is a diagonal matrix, positivity would hold if $\omega, a, b > 0$.

But those are strict restrictions.

Suppose that $h_i > 0$ $\mathbb{P}$ a.s. $\forall\, i < t$ $\omega > 0, a \geq 0, b \geq 0$

$$h_t = \omega + a y_{t-1}^2 + b h_{t-1} \geq \omega + b h_{t-1} \quad \text{then } \exists\, c \in [0,1) : c \leq b$$

$$\geq \omega + c h_{t-1} \geq \omega + c\omega + c^2 h_{t-2} \geq \omega + c\omega + c^2\omega + c^3 h_{t-3}$$

$$\geq \dots \geq \omega \underbrace{\sum_{i=0}^{\infty} c^i}_{\text{which exists since } c \in [0,1)}$$

So positivity holds also under weak positivity of a and b.

(If $a, b = 0$ the recursion is trivial and we have conditional comoskedasticity)

- Existence of a Unique Stationary and Ergodic Solution to the stochastic recursion (other solutions may also exist but they won't be stationary ergodic).

We will use a lemma which has some conditions (which are satisfied if $E(\log^+ z_0^2) < +\infty$ where $\log^+(x) = \begin{cases} \ln x, & x > 1 \\ 0, & x \leq 1 \end{cases}$, and we have that in our case $0 \leq E(\log^+ z_0^2) \leq E(z_0^2) = 1 < +\infty$ so the conditions hold)

$$h_t = \omega + (a z_{t-1}^2 + b) h_{t-1} = f(z_{t-1}, h_{t-1})$$

$$f(x,y) = \omega + (ax^2 + b)y$$

$$\frac{\partial f(z_0,y)}{\partial y} = a z_0^2 + b$$

$$\sup_{y>0} \left| \frac{\partial f(z_0,y)}{\partial y} \right| = \sup_{y>0} |a z_0^2 + b| = |a z_0^2 + b| = a z_0^2 + b$$

So the general lemma says that if $E \ln(a z_0^2 + b) < 0$ (or $-\infty$) then the GARCH(1,1) recursion admits a unique stationary and ergodic solution which is obtained as a "limit" of backward substitutions.

By Jensen's inequality $E \ln(a z_0^2 + b) \leq \ln E(a z_0^2 + b) = \ln(a+b)$

So if $a+b<1$ then $E\ln(a z_0^2 + b) < 0$ (but this does not go the other way)

It is possible that $E\ln(a z_0^2 + b) < 0$ even for some values of a, b : $a+b \geq 1$

e.g. $a>0$, $b=0$ (ARCH(1) case)

The condition is $E\ln(a z_0^2) < 0 \Leftrightarrow \ln a + E \ln z_0^2 < 0 \Leftrightarrow \ln a < -2E\ln|z_0|$

$\Leftrightarrow a < \exp(-2E\ln|z_0|)$

If $z_0 \sim N(0,1)$ then $\exp(-2E\ln|z_0|) \simeq 3{,}56 > 1$

So it is possible that $a>1$ and there exists a unique stationary ergodic solution (But we will see that weak stationarity does not hold)

When $a=b=0$ the condition trivially holds, the logarithm equals $-\infty$.

$$h_t = w + (a z_{t-1}^2 + b) h_{t-1} = w + w(a z_{t-1}^2 + b) + (a z_{t-1}^2 + b)(a z_{t-2}^2 + b) h_{t-2}$$

$$= w + w(a z_{t-1}^2 + b) + w(a z_{t-1}^2 + b)(a z_{t-2}^2 + b) + (a z_{t-1}^2 + b)(a z_{t-2}^2 + b)(a z_{t-3}^2 + b) h_{t-3}$$

$$= w\left[1 + \sum_{i=1}^{m-1} \prod_{j=1}^{i} (a z_{t-j}^2 + b)\right] + \prod_{j=1}^{m} (a z_{t-j}^2 + b) h_{t-m}$$

$$\ldots = \lim_{m\to\infty} w\left[1 + \sum_{i=1}^{m-1} \prod_{j=1}^{i} (a z_{t-j}^2 + b)\right] + \lim_{m\to\infty} \prod_{j=1}^{m} (a z_{t-j}^2 + b) c$$

where c is an appropriate positive constant.

$$= w\left[1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (a z_{t-j}^2 + b)\right] + (\qquad)$$

Let's assume that $-\infty \leq E\ln(az_0^2+b) < 0$

$$\prod_{j=1}^{m}(az_{t-j}^2+b) = \exp\left(\sum_{j=1}^{m}\ln(az_{t-j}^2+b)\right) = \exp\left(m\frac{1}{m}\sum_{j=1}^{m}\ln(az_{t-i}^2+b)\right)$$

$$\frac{1}{m}\sum_{j=1}^{m}\ln(az_{t-j}^2+b) \xrightarrow{m\to\infty} \underbrace{E\ln(az_0^2+b)}_{<0} \quad P \text{ a.s.}$$

So m times sth negative goes to $-\infty$ and the exponential goes to 0

i.e. $\exp\left(m\frac{1}{m}\sum_{j=1}^{m}\ln(az_{t-j}^2+b)\right) \to 0 \quad P$ a.s.

Then the unique stationary ergodic solution to the GARCH recursion is:

$$h_t = \omega\left[1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}(az_{t-j}^2+b)\right].$$

Is this comformablee to our framework?

Is it adopted to $\mathcal{F}_t$? Yes.

Positivity? The positivity conditions are verified, ω cannot be 0 since then $h_t=0$ and if $a\geq 0$, $b\geq 0$ then $h_t > 0$.

$$\text{When } a=0,\quad h_t=\omega\left[1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}b\right] = \omega\sum_{j=0}^{\infty}b^i = \frac{\omega}{1-b} \qquad \forall\, t\in\mathbb{Z}$$

conditionally homoskedastic.

Note that,

When $a\neq 0$, h_t is not a linear process.

So when $-\infty \leq E\ln(az_0^2+b) < 0$ (*), (h_t) exists and is stationary ergodic. This implies that the GARCH(1,1) process (y_t) is also stationary ergodic.

(*) Is a necessary condition but also sufficient.

Is (y_t) weakly stationary? Is the above sufficient?

$$E(h_t) = E\left[\omega\left[1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}(az_{t-j}^2+b)\right]\right]$$

$$= \omega\left[1+\sum_{i=1}^{\infty}E\left(\prod_{j=1}^{i}(az_{t-j}^2+b)\right)\right]$$

$$\overset{z \text{ indep.}}{=} \omega\left[1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left(a\,E(z_{t-j}^2)+b\right)\right] = \omega\left[1+\sum_{i=1}^{\infty}\prod_{j=1}^{i}(a+b)\right]$$

$$= \omega\left[1+\sum_{i=1}^{\infty}(a+b)^i\right] = \underbrace{\omega\sum_{i=0}^{\infty}(a+b)^i}_{\substack{\text{If } a+b<1 \\ \text{then it converges}}} = \begin{cases} +\infty, & a+b\geq 1 \\ \dfrac{\omega}{1-(a+b)}, & a+b<1 \end{cases}$$

So iff $a+b<1$ (remember that $a+b<1$ implies $-\infty < E\ln(a z_0^2+b)<0$) then (y_t) is strictly and weakly stationary and ergodic s.m.d. process.
But if $a+b\geq 1$ then (y_t) is stationary ergodic but not weakly stationary.

ARMA Representation for the squared process

When does $E(h_t^2)<+\infty$?

$$E(h_t^2) = E\left[\left(\omega + (a z_{t-1}^2+b)\,h_{t-1}\right)^2\right] = E\left[\omega^2 + 2\omega(a z_{t-1}^2+b)h_{t-1} + (a z_{t-1}^2+b)^2 h_{t-1}^2\right]$$

$$= E\left[\omega^2 + \left(a^2 z_{t-1}^4 + b^2 + 2a z_{t-1}^2 b\right)h_{t-1}^2 + 2\omega\left(a z_{t-1}^2+b\right)h_{t-1}\right]$$

$$\overset{z_t \text{ indep}}{=} \omega^2 + E(a^2 z_{t-1}^4 + b^2 + 2a z_{t-1}^2 b)\,E(h_{t-1}^2) + 2\omega\,\underbrace{E(a z_{t-1}^2+b)}_{=\,a+b}\,\underbrace{E(h_{t-1})}_{=\,\frac{\omega}{1-(a+b)}}$$

We have to assume that $E(z_0^4)\cdot u<+\infty$
(and by Jensen's inequality it can't be greater than 1)

$$= \omega^2 + (a^2 u + b^2 + 2ab)\,E(h_{t-1}^2) + 2\omega(a+b)\frac{\omega}{1-(a+b)}$$

$$= P(\omega,a,b) + A(a,b,u)\,E(h_{t-1}^2)$$

where $A(a,b,u) := E(a^2 z_{t-1}^4 + b^2 + 2a z_{t-1}^2 b) = a^2 u + b^2 + 2ab$

$P(\omega,a,b) := 2\omega\,E(a z_{t-1}^2+b)\,E(h_{t-1}) + \omega^2 = 2\omega(a+b)\dfrac{\omega}{1-(a+b)} + \omega^2$

iff $A(a,b,u) < 1$, $E(h_t^2) = \dfrac{P(\omega,a,b)}{1-A(a,b,u)}$

If the UDC holds, does the previous condition hold?

$a^2u + b^2 + 2ab \leq (a+b)^2$ since $u < 1$

< 1.

17/5/16

Example. GARCH(1,1) Process

- $\omega > 0$, $a \geq 0$, $b \geq 0$

$h_t = \omega + a y_{t-1}^2 + b h_{t-1} = \omega + (a z_{t-1}^2 + b) h_{t-1}$ (*)

- $E[\ln(a z_0^2 + b)] < 0$ or $-\infty$ is a necessary and sufficient condition for the existence of a unique strictly stationary and ergodic solution to (*).

(This condition is implied by the condition $a+b<1$ but the converse does not hold, e.g. ARCH(1) $(b=0)$, $z_0 \sim N(0,1)$ and $E(\ln a z_0^2) < 0 \Leftrightarrow a < 3{,}56$)

(We can show that there are also other solutions to the recursion (*) but they won't be strictly stationary)

- We have derived $h_t = \omega \left[1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (a z_{t-j}^2 + b)\right]$ non-linear process

- If $a+b<1 \Rightarrow E(h_t) = \dfrac{\omega}{1-(a+b)}$

- We have proved that iff $E(z_0^4) = u < +\infty$ $(u \geq 1)$ and $A(a,b,u) = a u^2 + b^2 + 2ab < 1$

(which is stronger than $a+b<1$)

(In the ARCH(1), $A(a,0,u) < 1 \Leftrightarrow a^2 u < 1 \Leftrightarrow a < \dfrac{1}{\sqrt{u}}$
and if $z_0 \sim N(0,1)$ i.e. $a < \dfrac{1}{\sqrt{3}}$ even stricter condition)

then $E(h_t^2) = \dfrac{P(\omega,a,b)}{1-A(a,b,u)}$ exists.

Lemma. Iff $\overbrace{E(z_0^4) < +\infty \text{ and } A(a,b,u) < 1}^{(**)}$ then the $(v_t)_{t\in\mathbb{Z}}$ process, $v_t := (z_t^2 - 1) h_t$, $t \in \mathbb{Z}$ is a strictly stationary and ergodic process.

Proof: v_t is a pointwise product of str. stationary and ergodic s.m.d. process

& $E(v_t^2) = E\left((z_t^2-1)^2 h_t^2\right) \overset{\text{indep.}}{=} \underbrace{E\left[(z_t^2-1)^2\right] E(h_t^2)}_{} = E(z_t^4 - 2 z_t^2 + 1) \dfrac{P(\omega,a,b)}{1-A(a,b,u)}$

$= (u-1) \dfrac{P(\omega,a,b)}{1-A(a,b,u)} < +\infty$

conditions (**) ensure the existence of those expectations

G $(v_t)_{t\in\mathbb{Z}}$ is adopted to $(G_t)_{t\in\mathbb{Z}}$, $G_t = \sigma(z_{t-i}, i \geq 0)$

by construction z_t^2 is adopted to G_{t-1} $\Rightarrow$ $z_t^2 - 1$ is adopted to G_t.

G $E(v_t / G_{t-1}) = E(v_t / f_t) = E\big((z_t^2-1)h_t / f_t\big) = h_t E(z_t^2 - 1 / f_t)$
$\overset{\text{ind.}}{=} h_t E(z_t^2 - 1) = 0$

<u>Lemma.</u> Iff $E(z_0^4) < +\infty$ and $A(a,b,u) < 1$ then the $(y_t^2)_{t\in\mathbb{Z}}$ is the solution of the following ARMA(1,1)

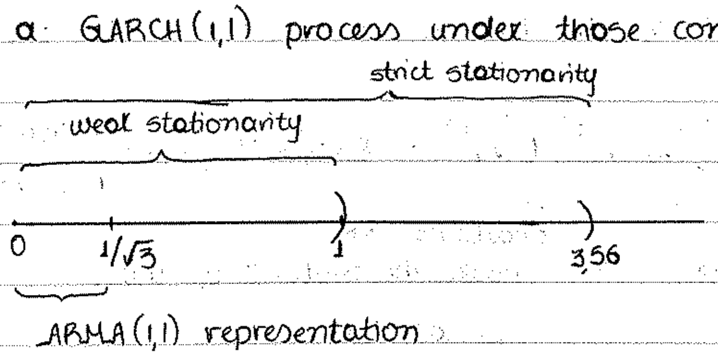
$$y_t^2 = \omega + (a+b)\, y_{t-1}^2 - b v_{t-1} + v_t, \quad t \in \mathbb{Z} \iff$$
$$y_t^2 - (a+b)\, y_{t-1}^2 = \omega - b v_{t-1} + v_t, \quad t \in \mathbb{Z}$$

Identification of the ARMA parameters
$\phi_1 = a+b$, $\theta_1 = b$
$\sigma^2 = (u-1)\dfrac{P(\omega,a,b)}{1 - A(a,b,u)}$

<u>Proof.</u> $h_t = \omega + a y_{t-1}^2 + b h_{t-1} \quad \forall\, t \in \mathbb{Z}$

$$\iff y_t^2 - h_t = y_t^2 - \omega - a y_{t-1}^2 - b h_{t-1} \quad \forall\, t\in\mathbb{Z}$$
$$\iff (z_t^2 - 1) h_t = y_t^2 - \omega - a y_{t-1}^2 - b h_{t-1} \quad \forall\, t\in\mathbb{Z}$$
$$\iff v_t = y_t^2 - \omega - a y_{t-1}^2 - b h_{t-1}, \quad \forall\, t\in\mathbb{Z}$$
$$\iff y_t^2 = \omega + a y_{t-1}^2 + b h_{t-1} + v_t \quad \forall\, t\in\mathbb{Z}$$
$$\iff y_t^2 = \omega + a y_{t-1}^2 \pm b y_{t-1}^2 + b h_{t-1} + v_t \quad \forall\, t\in\mathbb{Z}$$
$$\iff y_t^2 = \omega + (a+b) y_{t-1}^2 - b(y_{t-1}^2 - h_{t-1}) + v_t \quad \forall\, t\in\mathbb{Z}$$
$$\iff y_t^2 = \omega + (a+b) y_{t-1}^2 - b\big((z_t^2-1) h_{t-1}\big) + v_t \quad \forall\, t\in\mathbb{Z}$$
$$\iff y_t^2 = \omega + (a+b) y_{t-1}^2 - b v_{t-1} + v_t$$

So a GARCH(1,1) process under those conditions is simply an ARMA(1,1).



↳ In the ARCH(1) case the previous simplifies to AR(1) representation of the squares

↳ What if $a=0$, then we have an ARMA(1,1) with common root $\Rightarrow V_t$: WN process

E.g. ARCH(1) $(b=0)$ then $y_t^2 = \omega + a y_{t-1}^2 + V_t$

whence, if $\gamma_K^* := Cov(y_t^2, y_{t-K}^2) = \frac{b_1^K}{1-b_1^2} \sigma^2 = \frac{a^K}{1-a^2}(u-1)\frac{P(\omega,a,0)}{1-A(a,0,u)}$

<u>Some more indicative issues about conditionally heteroskedastic models</u>

▸ AR(1) - ARCH(1) process

Suppose that $(z_t)_{t\in \mathbb{Z}}$, iid, $E(z_0)=0$, $E(z_0^2)=1$, $E(z_0^3)=0$, $E(z_0^4)=u<+\infty$, $|b_0|<1$, $\omega>0$, $a\in[0, 1/\sqrt{u})$ and define the $(y_t)_{t\in\mathbb{Z}}$ as a solution of

$$\begin{cases} y_t = b_0 y_{t-1} + \epsilon_t \\ \epsilon_t = z_t h_t^{1/2} \\ h_t = \omega + a\epsilon_{t-1}^2 = \omega + a z_{t-1}^2 h_{t-1} \end{cases}, \quad t\in\mathbb{Z}$$

We have no positivity concerns, a satisfies strict stationarity conditions and weak stationarity, so ϵ_t exists and since UDC $|b_0|<1$ holds, y_t exists.
We are interested in inferencial procedures for b_0. The OLS estimator is:

$$b_T = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = b_0 + \frac{1/T \sum_{t=1}^T \epsilon_t y_{t-1}}{1/T \sum_{t=1}^T y_{t-1}^2}$$

Using all the arguments concerning the processes (ϵ_t) and (y_t) we can prove that $(\epsilon_t y_{t-1})$ is strictly stationary and ergodic and (y_{t-1}^2) is also stationary ergodic. So we may use the LLN of Birkoff.

Does $E(\epsilon_t y_{t-1})$ exist? It would exist if $E(|\epsilon_t y_{t-1}|)<+\infty$.
Remember that,

$$\left.\begin{aligned} y_t &= \sum_{i=0}^{\infty} b_0^i \epsilon_{t-i} \\ \epsilon_t &= \omega\left[1+\sum_{p=1}^{\infty}\prod_{m=1}^{p} a z_{t-m}^2\right] = \omega\left[1+\sum_{p=1}^{\infty} a^p \prod_{m=1}^{p} z_{t-m}^2\right] \end{aligned}\right\}$$ essentially this is the unique solution of the system above

By the Cauchy - Schwarz inequality:

$$E(|\epsilon_t y_{t-1}|) \leq \underbrace{\left(E(\epsilon_t^2)\right)^{1/2}}_{\substack{\text{condition for} \\ \text{ARCH holds} \\ \Rightarrow \text{expectation} \\ \text{exists}}} \underbrace{\left(E(y_{t-1}^2)\right)^{1/2}}_{\substack{\text{UDC holds} \\ \Rightarrow \text{expectation exist}}} < +\infty$$

Thereby, $E(\epsilon_t y_{t-1}) = E(z_t h_t^{1/2} y_{t-1}) \overset{LIE}{=} E\left[E(z_t h_t^{1/2} y_{t-1} / f_t)\right]$

$$= E\left(h_t^{1/2} y_{t-1} E(z_t / f_t)\right) = E\left(h_t^{1/2} y_{t-1} \overset{0}{E(z_t)}\right) = 0$$

So by Birkoff's LLN:

$$\frac{1}{T} \sum_{t=1}^{T} \epsilon_t y_{t-1} \longrightarrow E(\epsilon_t y_{t-1}) = 0 \quad a.s.$$

As we said before $E(y_{t-1}^2)$ exist and

$$E(y_{t-1}^2) = \frac{1}{1-b_0^2} E(\epsilon_0^2) = \frac{1}{1-b_0^2} E(h_t) = \frac{1}{1-b_0^2} \frac{\omega}{1-\alpha} > 0.$$

So, by Birkoff's LLN:

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2 \longrightarrow E(y_{t-1}^2) = \frac{\omega}{(1-b_0^2)(1-\alpha)} \quad a.s.$$

and since $E(y_{t-1}^2) > 0$ by CMT:

$$b_T \longrightarrow b_0 + \frac{0}{\frac{\omega}{(1-b_0^2)(1-\alpha)}} \quad a.s. \quad as \quad T \to \infty$$

$$b_T \to b_0 \quad a.s. \quad as \quad T \to \infty$$

Now, what about the asymptotic distribution of the OLSE.

Our results will be based upon the behavior of these two series

I. $\displaystyle\sum_{\substack{i,j=0 \\ i \neq j}}^{\infty} b_0^i b_0^j E(\epsilon_t^2 \epsilon_{t-1-i} \epsilon_{t-1-j})$

II. $\displaystyle\sum_{i=0}^{\infty} b_0^{2i} E(\epsilon_t^2 \epsilon_{t-1-i}^2)$

$$\text{II. } \sum_{i=0}^{\infty} \beta^{2i} E(\epsilon_t^2 \epsilon_{t-1-i}^2) = \sum_{i=0}^{\infty} \beta^{2i} \left(\gamma_{i+1}^* + \left(E(\epsilon_0^2)\right)^2 \right)$$

$$= \sum_{i=0}^{\infty} \beta^{2i} \left[\frac{\alpha^{i+1}}{1-\alpha^2} \frac{P(\omega,\alpha,0)}{1-A(\alpha,0,u)} (u-1) + \left(\frac{\omega}{1-\alpha}\right)^2 \right]$$

$$= \frac{\alpha}{1-\alpha^2} \frac{P(\omega,\alpha,0)}{1-A(\alpha,0,u)} (u-1) \underbrace{\sum_{i=0}^{\infty} (\underset{<1}{\beta_0^2} \underset{<1}{\alpha})^i}_{\text{converges}} + \left(\frac{\omega}{1-\alpha}\right)^2 \sum_{i=0}^{\infty} (\underset{<1}{\beta_0^2})^i$$

$$= \frac{\alpha}{1-\alpha^2} \frac{P(\omega,\alpha,0)}{1-A(\alpha,0,u)} (u-1) \frac{1}{1-\beta_0^2\alpha} + \left(\frac{\omega}{1-\alpha}\right)^2 \frac{1}{1-\beta_0^2}$$

$$= C(\omega, \alpha, u, \beta_0)$$

I. $\sum_{\substack{i,j=0 \\ i\neq j}}^{\infty} \beta_0^i \beta_0^j E(\epsilon_t^2 \epsilon_{t-1-i} \epsilon_{t-1-j})$ we can show that this expectation exist

if $E(|\epsilon_t^2 \epsilon_{t-1-i} \epsilon_{t-1-j}|)$ exists.

By Cauchy-Schwarz inequality

$$E(|\epsilon_t^2 \epsilon_{t-1-i} \epsilon_{t-1-j}|) \leq \left(E(\epsilon_t^4)\right)^{1/2} \left(E(\epsilon_{t-1-i}^2 \epsilon_{t-1-j}^2)\right)^{1/2}$$

$$\leq \left(E(\epsilon_t^4)\right)^{1/2} \left(E(\epsilon_{t-1-i}^4)\right)^{1/4} \left(E(\epsilon_{t-1-j}^4)\right)^{1/4}$$

$$\overset{\text{stat.}}{=} E(\epsilon_t^4) < +\infty.$$

$$E(\epsilon_t^2 \epsilon_{t-1-i} \epsilon_{t-1-j}) = E(z_t^2 h_t \epsilon_{t-1-i} \epsilon_{t-1-j}) \overset{LIE}{=} E\left(E(z_t^2 h_t \epsilon_{t-1-i} \epsilon_{t-1-j} / f_t)\right)$$

$$= E\left(h_t \epsilon_{t-1-i} \epsilon_{t-1-j} E(z_t^2 / f_t)\right) = E\left(h_t \epsilon_{t-1-i} \epsilon_{t-1-j} E(z_t^2)\right) = E(h_t \epsilon_{t-1-i} \epsilon_{t-1-j}$$

Suppose that $i < j$.

$$E(h_t \epsilon_{t-1-i} \epsilon_{t-1-j}) = E(h_t z_{t-1-i} h_{t-1-i}^{1/2} \epsilon_{t-1-j}) \overset{LIE}{=} E\left(E(h_t z_{t-1-i} h_{t-1-i}^{1/2} \epsilon_{t-1-j} / f_{t-1-i})\right)$$

$$= E\left(h_{t-1-i}^{1/2} \epsilon_{t-1-j} E(h_t z_{t-1-i} / f_{t-1-i})\right)$$

▸ $E(h_t z_{t-1-i} / f_{t-1-i})$

Remember that $h_t = \omega + \sum_{p=1}^{i} \alpha^p \prod_{m=1}^{p} z_{t-m}^2 + \alpha^{i+1} \prod_{m=1}^{i+1} z_{t-m}^2 h_{t-1-i}$

$$E(h_t z_{t-1-i} / f_{t-1-i}) = E(z_{t-1-i}\,\omega / f_{t-1-i}) + E\left(z_{t-1-i} \sum_{p=1}^{i} \alpha^p \prod_{m=1}^{p} z_{t-m}^2 / f_{t-1-i}\right)$$

$$+ E\left(z_{t-1-i}\, \alpha^{i+1} \prod_{m=1}^{i+1} z_{t-m}^2 h_{t-1-i} / f_{t-1-i}\right)$$

$$= \omega\, \cancelto{0}{E(z_{t-1-i})} + E\left(z_{t-1-i} \sum_{p=1}^{i} \alpha^p \prod_{m=1}^{p} z_{t-m}^2\right)$$

$$+ h_{t-1-i} E\left(z_{t-1-i}\, \alpha^{i+1} \prod_{m=1}^{i+1} z_{t-m}^2\right)$$

$$= \cancelto{0}{E(z_{t-1-i})}\, E\left(\sum_{p=1}^{i} \alpha^p \prod_{m=1}^{p} z_{t-m}^2\right) + h_{t-1-i} E\left(z_{t-1-i}^3\, \alpha^{i+1} \prod_{m=1}^{i} z_{t-m}^2\right)$$

$$= h_{t-1-i}\, \cancelto{0}{E(z_{t-1-i}^3)}\, \alpha^{i+1} E\left(\prod_{m=1}^{i} z_{t-m}^2\right) = 0$$

Therefore, $\sqrt{T}(b_T - b_0) \xrightarrow{d} N\left(0,\ C(\omega, \alpha, u, b_0)\, E(y_0^2)^{-2}\right)$ as $T \to \infty$.

18/5/16

Can we obtain a consistent estimator of the asymptotic variance?
If we had consistent estimators of the parameters ω, α, u, b_0, we would obtain a consist. estimator of the asy. variance. But, this would be a parametric/semi-parametric estimator. We need to assume some specification; take as given for example that we have a GARCH.
Can we obtain a non-parametric estimator?
Can we use sample moments? Something like $\dfrac{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 y_{t-1}^2}{\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^2}$ but we don't observe ϵ and this is an explicit function of the ϵs. Would it be possible somehow to estimate ϵ given the sample and the OLSE? We could use the residuals of the regression.
Define $e_t := y_t - b_T y_{t-1}$, $\forall t = 1, \dots, T$
and it could be possible to obtain a consistent estimator of the asymptotic variance of the OLSE as $V_T = \dfrac{\frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1}^2}{\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^2}$

The question is what are the properties of such an estimator?
It is a non-parametric estimator because it does not depend on parametric assumptions about the conditional variance.

We know that (y_{t-1}^2) is strictly stationary and ergodic and since $E(y_0^2)$ exists, using Birkoff's LLN and the CMT we know it's asymptotic limit.
What about the nominator?
Consider the difference:

$$e_t^2 y_{t-1}^2 - \epsilon_t^2 y_{t-1}^2 = (y_t - b y_{t-1})^2 y_{t-1}^2 - \epsilon_t^2 y_{t-1}^2 = (b_0 y_{t-1} + \epsilon_t - b_T y_{t-1})^2 y_{t-1}^2 - \epsilon_t^2 y_{t-1}^2 =$$

$$\left((b_0 - b_T) y_{t-1} + \epsilon_t\right)^2 y_{t-1}^2 - \epsilon_t^2 y_{t-1}^2 = (b_0 - b_T)^2 y_{t-1}^4 + \cancel{\epsilon_t^2 y_{t-1}^2} - 2(b_0 - b_T) y_{t-1}^3 \epsilon_t - \cancel{\epsilon_t^2 y_{t-1}^2} =$$

$$(b_0 - b_T)^2 y_{t-1}^4 - 2(b_0 - b_T) y_{t-1}^3 \epsilon_t$$

Essentially, we need to establish the asymptotic behavior of the distance:

$$\left| \frac{1}{T} \sum_{t=1}^{T} e_t^2 y_{t-1}^2 - E(\epsilon_0^2 y_{-1}^2) \right| =$$

$$\left| \frac{1}{T} \sum_{t=1}^{T} e_t^2 y_{t-1}^2 \pm \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2 y_{t-1}^2 - E(\epsilon_0^2 y_{-1}^2) \right| \leq \quad \text{(by the triangle inequality)}$$

$$\underbrace{\left| \frac{1}{T} \sum_{t=1}^{T} e_t^2 y_{t-1}^2 - \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2 y_{t-1}^2 \right|}_{(*)} + \underbrace{\left| \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2 y_{t-1}^2 - E(\epsilon_0^2 y_{-1}^2) \right|}$$

(*) so this is the reason why we constructed the

we already know that $\frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2 y_{t-1}^2 \to E(\epsilon_0^2 y_{-1}^2)$ a.s.

By the triangle inequality:

$$(*) \leq \frac{1}{T} \sum_{t=1}^{T} \left| e_t^2 y_{t-1}^2 - \epsilon_t^2 y_{t-1}^2 \right| = \frac{1}{T} \sum_{t=1}^{T} \left| (b_0 - b_T)^2 y_{t-1}^4 - 2(b_0 - b_T) y_{t-1}^3 \epsilon_t \right|$$

$$\leq \frac{1}{T} \sum_{t=1}^{T} 2 |b_0 - b_T| |y_{t-1}^3 \epsilon_t| + \frac{1}{T} \sum_{t=1}^{T} (b_0 - b_T)^2 y_{t-1}^4$$

$$= \frac{1}{T} 2 |b_0 - b_T| \sum_{t=1}^{T} |y_{t-1}^3 \epsilon_t| + \frac{1}{T} (b_0 - b_T)^2 \sum_{t=1}^{T} y_{t-1}^4$$

We have derived that $b_0 \to b_T$ a.s. so we just need to show that the above sums converge, so (*) will converge to 0.

Using the Cauchy-Schwarz inequality we can show that those sums converge (if those sums diverge, the above could still converge to 0 if we prove that the sums diverge at a slower rate)

Thereby, $V_T \to C(\omega, a, u, b_0) (E(y_0^2))^{-2}$ a.s. as $T \to \infty$.

Now, what is the limiting behavior of

$$\frac{1}{\sqrt{T}} \sqrt{T} (b_T - b_0) \xrightarrow{d} N(0,1)$$

hence,

$$T (b_T - b_0)^2 \frac{1}{V_T} \xrightarrow{d} \chi_1^2$$

$$T(b_T-b_0)^2\frac{1}{V_T} = T(b_T-b_0)^2 \frac{1}{\dfrac{\frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1}^2}{\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^2}} = T(b_T-b_0)^2 \frac{\frac{1}{T}\left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{\frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1}^2}$$

$$= (b_T-b_0)^2 \frac{\left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{\sum_{t=1}^{T} e_t^2 y_{t-1}^2} := W_T(b_0) \xrightarrow{d} \chi_1^2.$$

Given this we can construct a Wald type test for:

$(|b^*| < 1)$

$H_0: b_0 = b^*$

$H_1: b_0 \neq b^*$

Given a level of significance $(\alpha \in (0,1))$ the quantile of χ^2 is the inverse of the cdf:

In the particular distribution where the cdf is continuous, this is an equality.

$q_{\chi^2}(1-\alpha) = \inf\left(x \in \mathbb{R},\ P(u \leq x) \geq 1-\alpha\right) \quad u \sim \chi_1^2$

> 0 (we can prove it in this particular case)

So the test statistic is

$$W_T(b^*) = (b_T-b^*)^2 \frac{\left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{\sum_{t=1}^{T} e_t^2 y_{t-1}^2}$$

and what is the rejection region?

iff $W_T(b^*) > q_{\chi_1^2}(1-\alpha)$ we reject the null

Testing procedure: reject H_0 at α iff $W_T(b^*) > q_{\chi_1^2}(1-\alpha)$

- The asymptotic exactness of a test relies on the probability of rejecting the null when the null is true and asymptotically is equal to α
- The consistency of a test relies on the probability of rejecting the null when the null is false and it converges to 1.

<u>Proposition</u>: Given the asymption framework the testing procedure is:

a. asymptotically exact i.e. $\lim_{T\to\infty} P(W_T(b^*) > q_{\chi_1^2}(1-\alpha) / H_0) = \alpha$

b. consistent i.e. $\lim P(W_T(b^*) > q_{\chi_1^2}$

Proof: (a) Suppose that $\mathcal{H}_0$ is true, i.e. $W_T(\theta^*) = W_T(\theta_0) \xrightarrow{d} u \sim \chi^2_1$

So $P(W_T(\theta^*) > q_{\chi^2_1}(1-a) / \mathcal{H}_0) = P(W_T(\theta_0) > q_{\chi^2_1}(1-a))$

$$\downarrow T \to \infty$$

$P(u > q_{\chi^2_1}(1-a)) = a$ (by the definition of the quantile)

(b) Suppose that $\mathcal{H}_1$ is true, then $W_T(\theta^*) = (\theta_T - \theta^*)^2 \dfrac{\left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{\sum_{t=1}^{T} e_t^2 y_{t-1}^2} =$

$$T(\theta_T - \theta^*)^2 \frac{\frac{1}{T^2}\left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{\frac{1}{T}\sum_{t=1}^{T} e_t^2 y_{t-1}^2} = T(\theta_T - \theta_0 + \theta_0 - \theta^*)^2 \frac{1/T^2 \left(\sum_{t=1}^{T} y_{t-1}^2\right)^2}{1/T \sum_{t=1}^{T} e_t^2 y_{t-1}^2} =$$

$$\Big[\underbrace{T(\theta_T - \theta_0)^2}_{\xrightarrow{d}\ u} + \underbrace{T(\theta_0 - \theta^*)^2}_{\text{diverges at rate } T} + 2\underbrace{\sqrt{T}(\theta_T - \theta_0)}_{\xrightarrow{d}\ u} \underbrace{\sqrt{T}(\theta_0 - \theta^*)}_{\text{diverges at rate } \sqrt{T}}\Big] V_T^{-1}$$

since $(\theta_0 - \theta^*)^2$ is sth positive, it diverges to $+\infty$

This rate will dominate.

So $P(\underbrace{W_T(\theta^*)}_{\text{diverges to } +\infty} > \underbrace{q_{\chi^2_1}(1-a)}_{\text{positive real number}} / \mathcal{H}_1) = 1$

Extension to the GARCH(p,q)

Definition. Suppose that $p, q \in \mathbb{N}$, $\omega > 0$, $a_i \geq 0$, $i = 1, \dots, p$, $b_j \geq 0$, $j = 1, \dots, q$ then the GARCH(p,q) recursion is:

$$h_t = \omega + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{j=1}^{q} b_j h_{t-j}$$

$$= \omega + \sum_{i=1}^{p} a_i z_{t-i}^2 h_{t-i} + \sum_{j=1}^{q} b_j h_{t-j}$$

$$= \omega + \sum_{i=1}^{\max(p,q)} (a_i^* z_{t-i}^2 + b_i^*) h_{t-i}$$

where $a_i^* = \begin{cases} a_i & \text{if } i \leq p \\ 0 & \text{if } i > p \end{cases}$, $b_i^* = \begin{cases} b_i & \text{if } i \leq q \\ 0 & \text{if } i > q \end{cases}$

EGARCH (1,1)

The general framework is the same.

Definition: ω, γ, a, b then the $(y_t)_{t\in\mathbb{Z}}$ process is termed as EGARCH(1,1) process iff $(h_t)_{t\in\mathbb{Z}}$ satisfies

$$h_t = \exp\left(\omega + a\left(|z_{t-1}| - E|z_{t-1}|\right) + \gamma z_{t-1} + b\ln h_{t-1}\right), \quad t \in \mathbb{Z}$$

due to the exponential positivity is ensured, we don't need to restrict the parameters $\Rightarrow$ lower computational cost

$$\Leftrightarrow \ln h_t = \omega + a\left(|z_{t-1}| - E|z_{t-1}|\right) + \gamma z_{t-1} + b\ln h_{t-1}, \quad t \in \mathbb{Z}$$

If we define $u_t := a\left(|z_{t-1}| - E|z_{t-1}|\right) + \gamma z_{t-1}$

$$\Rightarrow \ln h_t = \omega + b\ln h_{t-1} + u_t \qquad (*)$$

$(u_t)_{t\in\mathbb{Z}}$ is an iid process since (z_t) is an iid process with $E(u_t) = 0$ and $V(u_t) < +\infty$.

So $(*)$ is an AR(1) recursion with iid errors and there exists a unique stationary ergodic solution of the AR(1) recursion iff $|b| < 1$ which implies that (h_t) will have a unique stationary ergodic solution iff $|b| < 1$, which implies that (y_t) has stationary ergodic solution iff $|b| < 1$.

Lemma. Iff $|b| < 1$ there exists a unique stationary ergodic solution to $(*)$ and thereby the EGARCH(1,1) process $(y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic. Furthermore, the particular solution

$$h_t = \exp\left(\frac{\omega}{1-b} + \sum_{i=0}^{\infty} b^i \left(a\left(|z_{t-1-i}| - E|z_{t-1-i}|\right) + \gamma z_{t-1-i}\right)\right) \quad \forall\, t \in \mathbb{Z}$$

since z_t is iid $\sum_{i=0}^{\infty} b^i a E|z_{t-1-i}| = \sum_{i=0}^{\infty} b^i a E|z_0| = \dfrac{aE|z_0|}{1-b}$

Then

$$h_t = \exp\left(\frac{\omega^*}{1-b} + \sum_{i=0}^{\infty} b^i \left(a|z_{t-1-i}| + \gamma z_{t-1-i}\right)\right), \quad t \in \mathbb{Z}$$

where $\omega^* = \omega - a E|z_0|$

$$h_t = \exp\left(\frac{\omega^*}{1-b}\right) \prod_{i=0}^{\infty} \exp\left[b^i\left(a|z_{t-1-i}| + \gamma z_{t-1-i}\right)\right]$$

$$E(h_t) = \exp\left(\frac{\omega^*}{1-b}\right) E\left[\prod_{i=0}^{\infty} \exp\left[b^i\left(a|z_{t-1-i}| + \gamma z_{t-1-i}\right)\right]\right]$$

$$\overset{z_t \text{ indep.}}{=} \exp\left(\frac{\omega^*}{1-b}\right) \prod_{i=0}^{\infty} E\left[\exp\left[b^i\left(a|z_{t-1-i}| + \gamma z_{t-1-i}\right)\right]\right]$$

$$\overset{z_t \text{ ident.}}{=} \exp\left(\frac{\omega^*}{1-b}\right) \prod_{i=0}^{\infty} E\left[\exp\left[b^i\left(a|z_0| + \gamma z_0\right)\right]\right]$$

In order to exist this expectation, we need to specify stricter restrictions on the properties of the distribution of z. We need to impose restrictions about the existence of exponential moments of z.
(If $z \sim N(0,1)$ then the expectation exists for any value of the parameters)

- We can estimate using QML
  In the GARCH(1,1) case we obtain a consistent and asymptotically normal Gaussian QMLE even if only conditions of strict stationarity holds (we don't need weak stationarity conditions).

24/5/16

Primitive Notions of Unit Roots Econometrics

Consider $y_t = \beta_{1_0} y_{t-1} + \epsilon_t$, $t \in \mathbb{Z}$, where $(\epsilon_t)_{t\in\mathbb{Z}}$ is a white noise process and $\beta_{1_0} = 1$.

The UDC for the existence of a linear causal solution over the integers is not satisfied.

One way to solve this is to define it not over the integers but over a subset of them, over for example the $\mathbb{N}$.

The aforementioned is called a unit root recursion.

In order to solve the recursion is to set an initial condition. So,

$y_t = \beta_{1_0} y_{t-1} + \epsilon_t$, $t \in \mathbb{N}$, $(\epsilon_t)_{t\in\mathbb{N}}$, $\beta_{1_0} = 1$

$y_0 = 0$

a solution of which is

$$(y_t)_{t\in\mathbb{N}}, \quad y_t = \begin{cases} 0, & t=0 \\ \sum_{i=1}^{t} \epsilon_t, & t>0 \end{cases}$$: standard random walk process

It is easy to see that

$E(y_t) = 0 \quad \forall\ t \geq 0$

$$V(y_t) = \begin{cases} 0, & t=0 \\ \sum_{i=1}^{t} V(\epsilon_i), & t>0 \end{cases} = \begin{cases} 0, & t=0 \\ \sigma^2 t, & t>0 \end{cases} = \sigma^2 t$$

The variance depends on t, so the marginals will depend on $t \Rightarrow$ neither strictly stationary nor weakly stat.

It is important to know how to test

$H_0: \beta_{1_0} = 1$

$H_1: \beta_{1_0} \in (-1, 1)$

One way to construct a testing procedure is to examine the asymptotic behavior of the OLSE in this structure.

Preparation

1. (Standard) Wienner Process (Brownian motion)

It is a continuous type process, as such it defines a probability measure over interval and it is analogous to a probability distribution on a function space equivallent to a standard normal distribution.

Definition: A standard Wiemer process is a continuous time process over $[0,1]$ ($W(r,\omega)$, for a given r it is a r.v. while for a given ω it is a real valued fnc on $[0,1]$ and for given r, ω is a real number)
denoted as W, defined on a complete probability space $(\Omega, f, \mathbb{P})$, that has the following properties:

1. $W(0) = 0$ $\mathbb{P}$ a.s.
2. $\forall\, t,s \in [0,1]$ $t \neq s$, $\forall\, h \in \mathbb{R}$, $t+h, s+h \in [0,1]$, $W(t) - W(s)$ is independent of $W(t+h) - W(s+h)$ and the distribution of $W(t) - W(s)$ equals the distribution of $W(t+h) - W(s+h)$
(It is essentially a stationarity assumption)
3. $\forall\, t,s \in [0,1]$ $(t \neq s)$, $W(t) - W(s) \sim \mathcal{N}(0, |t-s|)$

How do we know that this process exists? (Karhunen-Loeve Th$^{\underline{m}}$)

Proposition

1. W exists
2. $\forall\, t > 0$, $W(t) \sim \mathcal{N}(0,t)$
3. (We can choose a representation of the process such that its sample paths are continuous fncs)
There exists a version of W such that almost every sample path of the process is continuous ($W(\cdot, \omega) : [0,1] \to \mathbb{R}$ is continuous $\mathbb{P}$ a.s.)
4. $\mathbb{P}$ a.s. every path is almost everywhere non-differentiable.

If we have a continuous fnc over $[0,1]$ its Riemann integral exists.

Suppose that $f: \mathbb{R} \to \mathbb{R}$ continuous, then $f(W)(\cdot, \omega): [0,1] \to \mathbb{R}$ is a composition of continuous fncs, so it is a fnc with continuous paths. The above along with the property of Riemann integral implies that $\int_0^1 f(W)(r)\,dr$ is a well-defined random variable.

E.g. Suppose that $f: \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$
According to the previous discussion $\int_0^1 W^2(r)\,dr$ is a well-defined r.v. and we can prove that $P\left(\int_0^1 W^2(r)\,dr > 0\right) = 1$

2. Partial Sum Process and a Functional Central Limit Th^{m} (FCLT)

Definition. Consider $(\epsilon_t)_{t \in \mathbb{N}}$ and for $T \geq 1$ $(T \in \mathbb{N}^*)$, $r \in [0,1]$, define

$$W_T(r) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor rT \rfloor} \epsilon_i \quad \text{where } \lfloor x \rfloor := \text{the integer part of } x.$$

This is essentially a sequence of such processes (for any particular value of r and T, W is a random variable).
$W_T(\cdot)$ is called the partial sum process of $(\epsilon_t)_{t \in \mathbb{N}}$ over the $[0,1]$.

Let's say that $1 \leq t < T$ and $t-1 \leq rT < t \Rightarrow \frac{t-1}{T} \leq r < \frac{t}{T}$
$\Downarrow$
the integer part of rT is $t-1$.

Thereby, $W_T(r) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor rT \rfloor} \epsilon_i = \frac{1}{\sqrt{T}} \sum_{i=1}^{t-1} \epsilon_i = W_T\left(\frac{t-1}{T}\right) \quad \forall r \in \left[\frac{t-1}{T}, \frac{t}{T}\right]$

hence, when $t-1 \leq rT < t$, $W_T(r)$ remains constant.
and when $r = \frac{t}{T}$, $W_T\left(\frac{t}{T}\right) = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \epsilon_i$ it is a jump process.

$$\blacktriangleright \int_0^1 W_T^2(r)\,dr = \sum_{t=1}^{T} \int_{\frac{t-1}{T}}^{\frac{t}{T}} W_T^2(r)\,dr \;\underset{\text{so } W_T^2 \text{ is also continuous}}{\overset{W_T \text{ is contin.}}{=\!=}}\; \sum_{t=1}^{T} W_T^2\left(\frac{t-1}{T}\right) \int_{\frac{t-1}{T}}^{\frac{t}{T}} dr =$$

$$\sum_{t=1}^{T} W_T^2\left(\frac{t-1}{T}\right)\frac{1}{T} = \frac{1}{T}\sum_{t=1}^{T} W_T^2\left(\frac{t-1}{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor \frac{t-1}{\cancel{T}}\cancel{T}\rfloor}\epsilon_i\right)^2 =$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\left(\sum_{i=1}^{t-1}\epsilon_i\right)^2 = \frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\epsilon_i\right)^2$$

FCLT

If $(\epsilon_t)_{t\in\mathbb{N}}$ is a stationary ergodic and s.m.d process (wrt some appropriate filtration) then $(\sigma^2 := E(\epsilon_0^2) > 0)$

$\frac{1}{\sigma}W_T \xrightarrow{d} W$ (convergence in a function space)

as $T\to\infty$.

(This also implies fidis convergence and thus convergence of marginal distrib.)

So,

$$\frac{1}{\sigma}W_T(1) = \frac{1}{\sigma}\frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor T\rfloor}\epsilon_i = \frac{1}{\sigma}\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\epsilon_i \xrightarrow{d} W(1) \sim N(0,1)$$

So this is the CLT for stationary ergodic s.m.d. processes.

Functional CMT

Suppose that G is a functional defined on the function space whose W_T, W attain their values and that it is appropriatelly continuous, then $G\left(\frac{1}{\sigma}W_T\right) \xrightarrow{d} G(W)$

e.g. $G(f) = \sigma f$ continuous

$G\left(\frac{1}{\sigma}W_T\right) \xrightarrow{d} G(W) \Leftrightarrow \sigma\frac{1}{\sigma}W_T \xrightarrow{d} \sigma W \Leftrightarrow W_T \xrightarrow{d} \sigma W$

e.g. $G(f) = \sigma^2 f^2$ continuous

$G\left(\frac{1}{\sigma}W_T\right) \xrightarrow{d} G(W) \Leftrightarrow \sigma^2\left(\frac{1}{\sigma}W_T\right)^2 \xrightarrow{d} \sigma^2 W^2 \Leftrightarrow W_T^2 \xrightarrow{d} \sigma^2 W^2$

e.g. $G(f) = \int_0^1 \sigma^2 f^2(r)\,dr$, continuous

$G\left(\frac{1}{\sigma}W_T\right) \xrightarrow{d} G(W) \Leftrightarrow \int_0^1 \sigma^2\left(\frac{1}{\sigma}W_T(r)\right)^2 dr \xrightarrow{d} \int_0^1 \sigma^2\left(W(r)\right)^2 dr \Leftrightarrow$

$\int_0^1 W_T^2(r)\,dr \xrightarrow{d} \int_0^1 \sigma^2 W^2(r)\,dr$

Proposition. If $(\epsilon_t)_{t\in\mathbb{N}}$ is a stationary ergodic s.m.d. process then

$$\begin{pmatrix} \frac{1}{T}\sum_{i=1}^{T}\epsilon_i^2 \\ W_T^2(1) \\ \int_0^1 W_T^2(r)\,dr \end{pmatrix} \xrightarrow{d} \sigma^2 \begin{pmatrix} 1 \\ W^2(1) \\ \int_0^1 W^2(r)\,dr \end{pmatrix} \quad \text{as } T\to\infty.$$

Sketch of Proof: By Birkoff's LLN $\frac{1}{T}\sum_{i=1}^{T}\epsilon_i^2 \to E(\epsilon_i^2)=\sigma^2$ a.s. and we have already proved the other two results. By the Cramer-Wold device we can prove this joint distribution.

Remember that the OLSE : $b_T = \dfrac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = 1 + \dfrac{\sum_{t=1}^{T}\epsilon_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$

- $\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 = \frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\epsilon_i\right)^2 = \int_0^1 W_T^2(r)\,dr \xrightarrow{d} \sigma^2\int_0^1 W^2(r)\,dr$

$y_t = y_{t-1}+\epsilon_t \overset{\forall t\geq 1}{\Longrightarrow} y_t^2 = (y_{t-1}+\epsilon_t)^2 = y_{t-1}^2+\epsilon_t^2+2\epsilon_t y_{t-1} \Rightarrow$

$$\epsilon_t y_{t-1} = \frac{1}{2}\left(y_t^2 - y_{t-1}^2 - \epsilon_t^2\right)$$

$$\sum_{t=1}^{T}\epsilon_t y_{t-1} = \frac{1}{2}\left(\sum_{t=1}^{T} y_t^2 - \sum_{t=1}^{T} y_{t-1}^2 - \sum_{t=1}^{T}\epsilon_t^2\right) = \frac{1}{2}\left(y_T^2 - \cancelto{0}{y_0^2} - \sum_{t=1}^{T}\epsilon_t^2\right)$$

$$= \frac{1}{2}\left(\left(\sum_{i=1}^{T}\epsilon_i\right)^2 - \sum_{t=1}^{T}\epsilon_t^2\right)$$

- $\frac{1}{T}\sum_{t=1}^{T}\epsilon_t y_{t-1} = \frac{1}{2}\left(\frac{1}{T}\left(\sum_{i=1}^{T}\epsilon_i\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\right) = \frac{1}{2}\left(\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{[T]}\epsilon_i\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\right)$

$$= \frac{1}{2}\left(W_T^2(1) - \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\right) \xrightarrow{d} \frac{1}{2}\left(\sigma^2 W^2(1) - \sigma^2\right) = \frac{\sigma^2}{2}\left(W^2(1)-1\right)$$

by the previous proposition, using the CMT.

Thereby, by the proposition, jointly,

$$\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \epsilon_t y_{t-1} \\ \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{\sigma^2}{2}\left(W^2(1)-1\right) \\ \sigma^2 \int_0^1 W^2(r)\,dr \end{pmatrix} \quad \text{as } T\to\infty$$

Then, by the CMT (since $\sigma^2 \int_0^1 W^2(r)\,dr > 0 \quad \mathbb{P}$ a.s.)

$$T(\beta_T - 1) = \frac{\frac{1}{T}\sum_{t=1}^{T} \epsilon_t y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \frac{1}{2}\underbrace{\frac{W^2(1)-1}{\int_0^1 W^2(r)\,dr}}_{(*)}$$

↳ Is OLS consistent? Yes, it's super consistent, it converges to 1 at a faster rate than $\sqrt{T}$, it converges at rate T.

↳ The r.v. $(*)$ has well-defined distribution that is non-degenerate and does not depend on σ^2.

↳ OLS is asymptotically biased.

▸ How would we construct a testing procedure?

$H_0: \beta_{1_0} = 1$

$H_1: \beta_{1_0} \in (-1,1)$

We can use as a statistic:

↳ $T(\beta_T - 1)$

or ↳ $T^2(\beta_T - 1)^2 \longrightarrow (\quad)^2 \quad \Rightarrow$ find the quantile fnc of the square

It is an asymptotically exact procedure → Dickey-Fuller testing type.

$$df = T^2(\beta_T - 1)^2 \xrightarrow{d} \frac{1}{4}\left(\frac{W^2(1)-1}{\int_0^1 W^2(r)\,dr}\right)^2 \longrightarrow q(1-\alpha)$$

under the H_0 use this as a critical value

reject H_1 if $T^2(\beta_T^2 - 1) > q(1-\alpha)$

27/5/16

In the unit root framework and if $(\epsilon_t)_{t\in\mathbb{N}}$ is stationary ergodic s.m.d. process, we obtained the limit theory of the OLS estimator.

$$T(b_T - 1) \xrightarrow{d} \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W^2(r)\,dr} := g$$

We are interested in using the above limit theory in order to construct a unit root test.

i.e. $\mathcal{H}_0: b_{1_0} = 1$

$\quad \mathcal{H}_1: b_{1_0} \in (-1,1)$

Using the square of g as a test statistic we can construct a Dickey-Fuller type of testing procedure.

<u>Dickey-Fuller Type unit root testing procedure</u>

1. $df_T = T^2(b_T - 1)^2$
2. Due to the CMT under $\mathcal{H}_0$, $df_T \xrightarrow{d} g^2$
3. Given the above we can construct an approximate rejection region: if $a \in (0,1)$ then we can define
   $q_{g^2}(1-a) := \inf(x \in \mathbb{R} : \mathbb{P}(g^2 \leq x) \geq 1-a) > 0$

Essentially we would perform this testing procedure, using a Monte Carlo to approximate this unknown quantile fnc.

Then reject the $\mathcal{H}_0$ at a iff $df_T > q_{g^2}(1-a)$

What about the properties of such a testing procedure?

- Asymptotic exactness

$$P(df_T > q_{g^2}(1-a) \,/\, \text{under } \mathcal{H}_0) = P(T^2(b_T - 1)^2 > q_{g^2}(1-a) \,/\, \text{under } \mathcal{H}_0)$$

$$\downarrow T\to\infty$$

$$P(g^2 > q_{g^2}(1-a)) = a$$

- Consistency

We need a probability of rejecting H_0 under H_1 which is the UDC.

We have shown $\sqrt{T}(b_T - b_{1_0}) \xrightarrow{d} N(0, V)$

e.g. $(\epsilon_t)_{t\in\mathbb{Z}}$ is an ARCH(1) process, ..., $E(z_0^3)=0$, $E(z_0^4):=u<+\infty$, $a<\frac{1}{\sqrt{u}}$

then

$$\sqrt{T}(b_T - b_{1_0}) \xrightarrow{d} N(0, V(\omega, a, u, b_{1_0}))$$

Thereby, under H_1:

$$T(b_T - 1) = T(b_T - b_{1_0}) + T(b_{1_0} - 1)$$

$$= \underbrace{\sqrt{T}(\sqrt{T}(b_T - b_{1_0}))}_{\substack{\text{this may converge} \\ \text{to 0 or diverge to} \\ -\infty \text{ or } +\infty\text{, at rate } \sqrt{T}}} + \underbrace{T(b_{1_0} - 1)}_{\substack{\text{this diverges to } +\infty \\ \text{at rate } T}}$$

$\Uparrow$

this has faster rate

$\Rightarrow$ it dominates

Then under H_1: df_T diverges to $+\infty$

$$P(df_T > q_{g^2}(1-a) \,/\, \text{under } H_1)$$

$$\downarrow T\to\infty$$

$$P(q_{g^2}(1-a) < +\infty) = 1$$ since it is a well-defined real number.

ATTENTION

These lecture notes are totally informal hence, they may contain several mistakes (of any type).

They do not substitute the lectures or the official notes of the course.

Please report any typos or mistakes to xatzilenan@aueb.gr