Financial Econometrics

Sol Yates

February 12, 2024

Contents

Lecture 2: Second Lecture - Review TS

Fri 02 Feb 15:45

1 Time Series Basics

Stochastic Processes

- chronologically ordered equidistant observations
- generated by stochastic process
- stochastic process - collection of RV (each Y_i is generated by different member of stochastic processes)
- **assumption** time series data has been generated by *stochastic process*

> **Definition 1.** **stochastic process** is a family of random variables defined on a probability space

> **Definition 2.** **time series** is a realisation of a stochastic processs

> **Definition 3.** **time series analysis** - only one history $Y_t(w)$, one state of the world $w \in \omega$ is available, but the goal is to derive the properties of $Y_t(\cdot)$ for a given t for different states of the world

idea - how can we understand what is driving omegas? Different states of the world, since we observe y_t. So place some structure on y_t

Should be able to recognise :

- non-stationary time series

- autoregressive time series
- kurtosis time series

Definition 4. auto covariance
Definition

time series often show correlation between successive observations, this feature is called serial correlating or **autocorrelation**

dependencies over time are described by auto covariance and autocorrelation functions

the j-th autocovariance of Y_t is given by

$$Cov[Y_t, y_{t-j}] = \gamma_{t,t-j} = E[Y_t - E[Y_t][Y_{t-j} - E[Y_{t-j}]$$

correspondingly the variance of Y_t is defined as :

$$V[Y_T] = \gamma_{t,t} = E[Y_t - E[Y_t])^2]$$

Definition 5. autocorrelation

the j-th autocorrelation of Y_t is given by :

$$\rho_{t,t-j} = \frac{Cov[Y_t, Y_{t-j}]}{V[Y_t]^{(\frac{1}{2})} V[Y_{t-j}^{(\frac{1}{2})}}$$

Definition 6. Covariance stationary A time series $\{Y_t\}_{t=-\infty}^{(\infty)}$ is called covariance stationary, or weakly stationary, if :

$$E[Y_t] = \mu_Y$$
$$V[Y_t] = \gamma_{t,t} = \gamma_0 = \sigma_Y^2 < \infty$$
$$Cov[Y_t, Y_{t-j}] = \gamma_{t,t-j} = \gamma_j < \infty$$

for a covaraince stationary process the j-th autocorrelation is given by :

Definition 7. white noise a TS is called this if it satisfies the following

$$E[Y_t] = 0 V[Y_t] = \sigma_Y^2 COv[Y_t, Y_s] = E[Y_t, Y_s] = 0$$

white noise is a weakly stationary process

Definition 8. Autocorrelation function of a covaraince stationary process $\{Y_t\}_{t=-\infty}^{(\infty)}$ is the sequence of autocorrelations ρ_j for all $j = 0, 1, 2, \ldots$

Definition 9. the empirical (or sample) autocorrelatino function of a time series Y_t is the sequence of sample autocorrelation coefficients $\hat{\rho}_j$ for all $j = 0, 1, 2, \ldots$:

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} = \frac{\sum_{t=j+1}^{T}(Y_t - \overline{Y}(Y_{t-j} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y}^2}$$

and

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \overline{Y})(Y_{t-j} - \overline{Y}) \qquad \overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

the graphical depictions of the empirical autocorrelation function is called an autocorrelogram

Definition 10. partial autocorrelation function

partial autocorrelation between Y_t and Y_{t-j} is the conditional correlation between Y_t and Y_{t-j} given (holding fixed) $Y_{t-1}, \ldots, Y_{t-j+1}$

$$a_j = Cor[Y_t, Y_{t-j}|Y_{t-1}, \ldots, Y_{t-j+1}]$$

effects of in-between values are controlled for

corresponding sample quantity $\hat{a}_j$ is called sample partial autocorrelation and is obtained as the OLS estimator of the coefficient a_j in model

$$Y_t = a_0 + a_1 Y_{t-1} + \ldots + a_j Y_{t-j} + \mu_t$$

Definition 11. sample autocorrelation function

if data generating process is a white noise process, then for large T:

$$\hat{\rho}_j \approx N(0, \frac{1}{T}), j = 1, 2, \ldots$$

Means : $H_0 : \hat{\rho}_j = 0$ is rejected, if zero does not fall within the approximate 95% confidence interval

$$[r\hat{h}o_j - \frac{2}{\sqrt{T}}, r\hat{h}o_j + \frac{2}{\sqrt{T}}]$$

Equivalently, autocorrelations are not significant when $\hat{\rho}_j$ is within the approximate two standard error bound $\pm 2/\sqrt{T}$

2 Arma Processes

Definition 12. a time series is called an autoregressive process of order p if it satisfies a relationship of the type :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \rho_p Y_{t-p} + \varepsilon_t$$

where ε_t is a white noise error term

A(1) process : the simplest form of an A(p) process is obtained for $p = 1$ as

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

Definition 13. MA(q)-Process a time series is called a **moving average process of order** q if it satisfies a relatinoship of the type

$$Y_t = \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

where ε_t is a white noise error term

MA(1) Process: the simplest form of an MA(q) process os obtained for $q = 1$ as

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Example. AR(1) Process

$$\sum_{i=0}^{\infty} \rho^{(i)} u_{t-i} =^{(wald)} MA(\infty)$$

2.1 ARMA

Lag operator let $\{Y_t\}_{t=-\infty}^{(\infty)}$ be a time series, then the lag operator $\mathcal{L}$ is defined by the relation

$$L^{(J)} \equiv Y_{t-j}$$

If $\{Y_t = c\}_{t=-\infty}^{(\infty)}$ where $c \in \mathbb{R}$, then $\mathcal{L}^{(j)} Y_t = L^{(j)} c = c$

ARMA(p,q) is a time series $\{Y_t\}_{t=-\infty}^{(\infty)}$ of the following form

$$\begin{aligned} \phi_p(L) Y_t &= c + \Theta(L) \varepsilon_t where \\ \phi_p(L) &= 1 - \phi_1 L - \phi_2 L^{(2)} - \ldots . - \phi_p L^{(p)} \\ \Theta(L) &= 1 + \theta_1 L + \theta_2 L^{(2)} + \ldots + \theta_q L^{(q)} \end{aligned}$$

with ε_t being a white noise and ϕ_p and Θ_q are called lag polynomials

2.2 ARMA estimation

ARMA(p, q) process:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 e_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

- estimation via conditional max likelihood
- **conditional** : derive the likelihood function under the assumption that the initial values of Y_t and ε_t are available
- **assume** : $\varepsilon_t \sim^{(iid)} N(0, \sigma^2)$
- ML parameter estimators are derided under the assumption of normality are quasi ML estimators

- our goal is to estimate the vector $\theta = (c, \phi_1, \phi_2, \ldots, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2)'$

ARMA estimation
Conditional log likelihood
estimation is done under asssumption that error term is normal.
LBQ test
whether p is sufficiently long, if model specified correctly, then residuals shouldn't be correlated with each other.
Tells whether white noise property is plausible assumption
critical values is from chi-squared dist, we test for absence of autocorrelation upto chosen lag order, leading into next weeks lecture of conditional heteroskedacity.
ARCH-LM test
tests for conditional heteroskecadity in regression residuals
pick ARMA based on this, if modelled succesfully then null of LBQ test shouldnt be rejected and there shouldnt be any conditional heterskedacity

Lecture 3: Univariate GARCH Models

Mon 12 Feb 09:01

Review

Week 1
Leptokurtic Property - How to measure a lot of outliers? Kurtosis. The kurtosis of our distribution is larger than 3 (4th moment of distribution). Since $K[r_t] > 3$ where $e \sim \mathcal{N}(\mu, \sigma^2)$
Left-Skewness - more negative returns than positive ones. $S[r_t] < 0$
Volatility clustering - periods of high volatility are followed by periods of high volatility. The volatile periods on the markets (across S of return distribution) they *cluster*. Market volatility is persistent.
Shape of daily returns - Compared to say a normal bell curve, is this a good distribution? Weekly more normal then daily, monthly more normal than weekly. Thus *aggregate returns tend to normality*

- Should know these by heart
- And be able to apply them and tell graphically

Week 2
Time series analysis
ARMA Models ($ARMA(1,1)$ $y_t = c + \rho_1 + y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$
Stationarity ADF test
Model selection ACF /DACF
Information criteria - Bayesian information criteria helps to choose whether ARMA(1, 2) or MA(1) is better for data.
At the end we estimate by quasi-likelihood since $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, which is key to today's material.
It is important to realise this assumption is quite strong, the shortcut for this type of estimator is quasi-likelihood.
$\Theta = (c, \rho_1, \theta_1, \sigma^2)$, in empirical terms the maximum likelihood estimators minimise the negative log-likelihood, we can only find the minimum using gradient descent, hence minimising the negative.
$\hat{\theta}_{QML} = \underset{\Theta}{argmin}$
Autoregressive process order 1
Log likelihood - maximises function to find combination of parameters of model such that our ε_t's are normal

For Financial Econometrics, once plot ACF and PCF, once looking at squared residuals, we have seen a lot of significant lags in the squared residuals. LBQ test and ARCH-LM test whether there is remaining autocorrelation within the squares residuals.

These tests tell us that $\hat{\sigma^2}$ tell us there is autocorrelation across time within the residuals, only problem of model misspecification comes from squared residuals, variance of error term.

2.3 Conditional Heteroskedacity

In any ARMA model there is some expectation

$$y_t = E[y_t|F_{t-1}] + \varepsilon_t$$

$c + \rho y_{t-1} + \theta\varepsilon_{t-1}$. F is filtration, past information and ε_t is new information/shock today.

White noise (DSA) :

$$\begin{aligned} \varepsilon_t &\sim WN \\ E[\varepsilon_t] &= 0 \\ V[\varepsilon_t] &= \sigma^2 \\ cov(\varepsilon_t, \varepsilon_t) &= 0 \end{aligned}$$

What is the difference between conditional and unconditional moments?

Conditional : $V[\varepsilon_t]$ and Unconditional $V[\varepsilon_t|F_{t-1}]$

$$\begin{aligned} y_t &= c + \rho y_{t-1} + \varepsilon_t \\ E[y_t] &= E[c + \rho y_{t-1} + \varepsilon_t] \\ &= c + \rho E[y_{t-1}] \\ E[y_t] &= E[y_{t-1}] \\ E[y_t] &= \frac{c}{1-\rho} \end{aligned}$$

That is,

$$\frac{c}{1-\rho} \quad vs \quad c + \phi y_{t-1} \qquad (*)$$

$$\begin{aligned} &E[y_t|F_{t-1}] \\ &E[c + \rho y_{t-1} + \varepsilon_t|F_{t-1}] \\ &c + \rho E[y_{t-1}|F_{t-1}] + E[\varepsilon_t|F_{t-1}] \\ &c + \rho y_{t-1} + 0 \end{aligned}$$

White Noise

- $E[\varepsilon_t] = 0$
- $V[\varepsilon_t] = \sigma^2$
- $cov[\varepsilon_t, \varepsilon_t] = 0$

The unconditional moment in (*) is more important.

White noise assumption, assumes both conditional and unconditional are constant over time, that is

$$V[\varepsilon_t] = V[\varepsilon_t, |F_{t-1}] = \sigma^2$$

$V[\varepsilon_t] = \sigma^2$ but $V[\varepsilon_t, F_{t-1}]$ is time varying (conditional second moment).

We start with $\varepsilon_t = \mathcal{L}_t \cdot \sigma_t$ where $\mathcal{L}_t \sim \mathcal{N}(0,1)$ and ARCH (1) : $\sigma_t^2 = w + \alpha\varepsilon_{t-1}$

As we have just done with AR1, now look at conditional and unconditional second moment of ARCH(1).

$V[\varepsilon_t]$ and

$$\begin{aligned} E[\varepsilon_t] = E[\mathcal{L}_t\sigma_t] &= \\ E[\mathcal{L}_t]E[\sigma_t] & \\ 0 \cdot E[\sigma_t] = 0 & \\ V[\varepsilon_t] = E[\varepsilon_t]^2 E[\mathcal{L}_t^2 \cdot \sigma_t^2] &= \\ E[\mathcal{L}_t^2] \cdot E[\sigma_t^2] = E[\sigma_t^2] & \end{aligned}$$

$V[\varepsilon_t|F_{t-1}]$ (not right yet)

$$\begin{aligned} E[\varepsilon_t|F_{t-1}] &= \\ E[\mathcal{L}_t]E[\sigma_t] & \\ 0 \cdot E[\sigma_t] = 0 & \\ V[\varepsilon_t] = E[\varepsilon_t]^2 E[\mathcal{L}_t^2 \cdot \sigma_t^2] &= \\ E[\mathcal{L}_t^2] \cdot E[\sigma_t^2] = E[\sigma_t^2] & \end{aligned}$$

General Settings

So far we have focused on the estimation of the conditional mean function $E[Y_t|F_{t-1}]$:

$$Y_t = E[Y_t|F_{t-1}] + \varepsilon_t$$

where ε_t is a weak white noise, that is, ε_t is serially uncorrelated : $Cov[\varepsilon_t, \varepsilon_{t-j}] = 0 \;\; \forall j \neq 1$

ARCH(1) Processes

A process σ_t^2 is called an ARCH(1) process if

$$\sigma_t^2 + w + \alpha\varepsilon_{t-1}^2$$

with w>0 and $\alpha \geq 0$

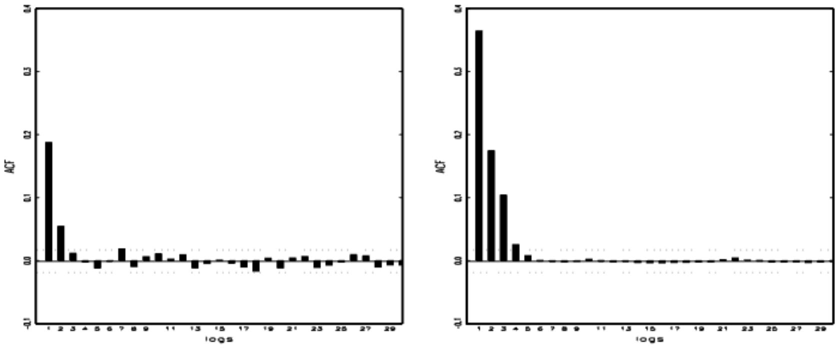
Properties of Arch(1)

- Arch (!) Conditional variance σ_t^2 is strictly positive if $w > 0$ and $a \geq 0$
- Opposite to the historical volatility estimator, the arch 1 volatility is a weighted average of past information that gives more weight to the recent information than to the distant one
- Arch 1 process can be written as an A(1) process in ε_t^2
- Consequently ε_t^2 is stationary if $|\alpha| < 1$

- Given that both process ε_t and ε_t^2 and $E[\varepsilon_t] = 0$ then the unconditional variance of ε_t, $E[\varepsilon_t]$ is given by

$$\sigma_\varepsilon^2 = V[\varepsilon_t] = E[\varepsilon_t^2] = \frac{w}{1-\alpha}$$

- ARCH(1) captures the clustering effect : when volatility is high, it more probably stays high
- The kurtosis is always large t



utocorrelation functions of squared time series with ARCH(1) conditional variance with $\alpha = 0.2$ (left panel) and $\alpha = 0.7$ (right panel)

Figure 1