Financial Econometrics: Long Memory and ARCH Effect

Shin-Huei Wang

CORE, Universite catholique de Louvain, Belgium

What is a time series?

[1.] It is an ordered sequence of data, where the order refers to increasing dates: $z_1, z_2, \ldots, z_n$ also denoted shortly $\{z_t\}_{t=1}^{n}$.

[2.] (2).The index t refers to a date: - For data without a time dimension (like a stock or a price) it is the exact moment (year, month, day, minute, hour, second...) when it is measured or observed. - For flow data, t refers to the period over which the flow variable is defined (January 2006, year 1925, third quarter of 1990, ...) or to a conventional date defining the end of the period.

[3.] Usually the dates are regularly spaced in time, but sometimes not: e.g. because of missing observations, or because the data are generated at random dates.

Stationary time series

(1.) What are ARMA models? What are the stationarity conditions?

(2.) For what series are they useful in practice, and for what purposes?

(3.) What method of estimation is available for these models?

(4.) How can we forecast the future using these models?

(I.)**Autoregressive models**

Definition of AR process

Autoregressive process of order p, AR(p) in brief:

$$y_t = \alpha_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

, where ϵ_t is $\sim (0, \sigma^2)$.

Stationarity of AR process

The AR(p) process is CS if $\phi_p(L)$ is stable.

(1) AR(1): the stability condition is $|\phi_1| < 1$, so that after substitutions

$$y_t = \alpha_0 \sum_{i=0}^{\infty} \phi_1^i + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} = \alpha_0 \sum_{i=0}^{\infty} \phi_1^i + \sum_{i=0}^{\infty} \phi_1^i L^i \epsilon_t$$

$$= \frac{\alpha_0}{1-\phi_1} + \frac{\epsilon_t}{1-\phi_1 L}$$

Stationarity corresponds to the fact that the impact multipliers (ϕ_1^i) of shocks tend to 0 quickly enough.

Stationarity of AR process

AR(p) process:

$$\phi_p(L) y_t = \alpha_0 + \epsilon_t$$

$$\Rightarrow y_t = \frac{\alpha_0}{\phi_p(L)} + \frac{\epsilon_t}{\phi_p(L)} = \frac{\alpha_0}{\phi_p(1)} + \frac{\epsilon_t}{\phi_p(L)}$$

$$= \quad \mu \quad + \psi(L)\epsilon_t,$$

where $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i = 1/\phi_p(L)$. The impact multipliers of shocks are the ψ_i coefficients.

From the last equality, $\mathrm{E}(\mathrm{y_t}) = \mu < \infty$ (by stability). Also, we can write the model as $\phi_p(L)(y_t - \mu) = \epsilon_t$.

(II.) Moving average models

Definition of MA process

A moving average process of order q, MA(q) in brief, is defined by

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

where $\epsilon_t \sim (0, \sigma^2)$.

With the lag polynomial notations, the MA(q) process is written

$$y_t = \mu + \theta_q(L)\epsilon_t, \quad \text{with}$$

$$\theta_q(L) = \theta_0 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q,$$

and with $\theta_0 = 1$.

Stationarity of MA process

MA(1): we see that $\mathrm{Var(y_t)} = \sigma^2(1+\theta_1^2)$, $\rho_1 = \theta_1 1 + \theta_1^2$, $\rho_j = 0$ for $j \geq 2$.

Also, $\mathrm{E(y_t)} = \mu$: the process is CS without restricting θ_1.

A MA(q) is CS without any restriction on the lag polynomial $\theta_q(L)$

(NO need for stability), with

$$\mathrm{E(y_t)} = \mu$$

$$\mathrm{Var(y_t)} = \sigma^2 \sum_{\mathrm{i=0}}^{\mathrm{q}} \theta_{\mathrm{i}}^2$$

$$(y_t, y_{t-j}) = \begin{cases} \sigma^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} & \text{if j} \leq q \\ \\ 0 & \text{if j ¿ q} \end{cases}$$

[III.] ARMA models

Definition of ARMA process

(1). Combining an AR(p) and a MA(q) in the same equation gives an ARMA(p,q) model:

$$y_t = \alpha_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$$

$$+ \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

where $\epsilon_t \sim (0,\sigma^2)$.

(2).

With the lag polynomial notations:

$$\phi_p(L) y_t = \underbrace{\alpha_0}_{\phi_p(1)\mu} + \theta_q(L)\epsilon_t.$$

Stationarity of ARMA process

(1). The stationarity conditions are exactly the same as for the AR(p) process. Then, we can write

$$\phi_p(L)y_t = \alpha_0 + \theta_q(L)\epsilon_t$$

$$y_t = \frac{\alpha_0}{\phi_p(1)} + \frac{\theta_q(L)}{\phi_p(L)}\epsilon_t$$

$$= \mu + \psi(L)\epsilon_t$$

$$\phi_p(L)(y_t - \mu) = \theta_q(L)\epsilon_t,$$

where $\psi(L) = \theta_q(L)/\phi_p(L) = \sum_{i=0}^{\infty} \psi_i L^i$.

(2). Stationarity corresponds to the fact that the impact multipliers (ψ_i) of shocks tend to 0 quickly enough if $i \to \infty$.

ARMA(1,1) process

$$y_t = \alpha_0 + \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$$

(1). The stationarity condition is $|\phi_1| < 1$. The MA form is

$$y_t = \frac{\alpha_0}{1-\phi_1} + \frac{1+\theta_1 L}{1-\phi_1 L}\epsilon_t$$

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

$$= \mu + \epsilon_t + (\theta_1 + \phi_1)\epsilon_{t-1} + \phi_1(\theta_1 + \phi_1)\epsilon_{t-2} + \dots$$

$$= \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

, so: $\psi_0 = 1$, $\psi_1 = \theta_1 + \phi_1$, $\psi_i = \phi_1^{i-1}(\theta_1 + \phi_1) = \phi_1 \psi_{i-1}$ for $i \geq 2$.

Estimation Methods

(1). AR models can be estimated by OLS. An alternative is ML. A version of ML (called conditional ML) is equivalent to OLS if one assumes the normality of error terms.

M odels having a MA part cannot be estimated by OLS, since past error terms are not observed: e.g. in $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$, we do not have observations on ϵ_{t-1}. For these models, conditional or exact ML estimation is available. An alternative is nonlinear least squares, which is the same as conditional ML assuming the normality of ϵ_t.

Invertibility of MA/ARMA processes

(1.) For stationarity of MA and ARMA processes, we have seen that no condition must be imposed on the MA polynomial. But for ML estimation, we must impose that this polynomial is stable: all roots of $\theta_q(z) = 0$ must exceed 1 in absolute value. Then we say that the MA or ARMA process is invertible.

(2.) For the MA(1) case, the condition is just that $|\theta_1| < 1$. The invertibility of $1 + \theta_1 L$ means that when we express ϵ_t as a function of y_t, y_{t-1}, ... and ϵ_0 as

$$\epsilon_t = (y_t - \mu) - \theta_1(y_{t-1} - \mu) + \theta_1^2(y_{t-2} - \mu) - \ldots + (-1)^{t-1}\theta_1^{t-1}(y_1 - \mu) + (-1)^t\theta_1^t\epsilon_0.$$

the coefficients θ_1^i tend to 0, instead of exploding if $|\theta_1| > 1$ (or being all 1 if $|\theta_1| = 1$).

(3.) For MA(q) or ARMA(p,q), invertibility means also that when we express ϵ_t as a function of y_t and lags of y_t, the coefficients tend to 0 when the lag order tends to ∞.

Is invertibility a restriction?

(1). Requiring invertibility is not really restrictive: any non invertible MA process can be transformed into an invertible one that has the same mean, variance, and autocorrelations. We illustrate this for the MA(1).

(2). Let $y_t = \mu + \tilde{\epsilon}_t + \tilde{\theta}_1 \tilde{\epsilon}_{t-1}$ with $\mathrm{Var}(\tilde{\epsilon}_\mathrm{t}) = \tilde{\sigma}^2$ and $|\tilde{\theta}_1| > 1$ (a non-invertible process). We know that $\mathrm{E}(\mathrm{y_t}) = \mu$, $\mathrm{Var}(\mathrm{y_t}) = \tilde{\sigma}^2(1 + \tilde{\theta}_1^2)$, $\rho_1 = \tilde{\theta}_1/(1 + \tilde{\theta}_1^2)$ and $\rho_j = 0$ for $j \geq 2$.

(3). Let $\theta_1 = 1/\tilde{\theta}_1$, so that $|\theta_1| < 1$, and $\epsilon_t = \tilde{\theta}_1 \tilde{\epsilon}_t$, so that $\mathrm{Var}(\epsilon_\mathrm{t}) = \sigma^2 = \tilde{\theta}_1^2 \tilde{\sigma}^2$. Next, define $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$, which is an invertible process. This process has the same mean, variance and autocorrelations as the non-invertible one above since $\tilde{\sigma}^2(1 + \tilde{\theta}_1^2) = \sigma^2(1 + \theta_1^2)$ and $\tilde{\theta}_1/(1 + \tilde{\theta}_1^2) = \theta_1/(1 + \theta_1^2)$.

I(1), I(0) and all that

When $y_t = \delta + y_{t-1} + u_t$ and u_t is stationary, y_t is also said to be "integrated of order 1", in short $y_t \sim I(1)$, because it is the sum (or integral) of a stationary process (plus a linear trend if $\delta \neq 0$):

$$y_t = y_0 + \delta t + \sum_1^t u_i.$$

Obviously, a RW corresponds to the particular case where u_t is an i.i.d. process. If $y_t \sim I((1), \Delta$is said to be I(0), so an I(0) process is stationary. So, we can say that a stationary process is "I(0)", with one exception: an over-differenced stationary process is said to be "I(-1)", even though it also stationary.

Economic implications

(1.) If y_t is a UR process, a shock affecting y today will be still felt many periods ahead.

(2.) For example, if log(GDP) is DS, a negative shock (e.g. oil shock, a big natural disaster, a war, a wrong economic policy decision...) will permanently decrease it by some amount. If nothing else happens in the future, GDP will be permanently below the level it would reach if no negative shock had occurred.

(3.) On the contrary, if y_t is stationary, the impact of a shock will vanish after a few periods: y_t will revert to its mean. For example, if log(GDP) is TS, a negative shock will decrease it in the future, but if nothing else happens in the future, it will come back to its trend level after a few quarters

Martingale

A martingale is a stochastic process such that the conditional expected value of an observation at some time t, given all the observations up to some earlier time s, is equal to the observation at that earlier time s. A sequence $Y_1, Y_2, Y_3, \ldots$ is said to be a martingale with respect to another sequence $X_1, X_2, X_3, \ldots$ if for all t.

$$E(|Y_t|) < \infty$$

and

$$E(Y_{t+1}|X_t, X_{t-1}, \ldots) = Y_t$$

Example: Suppose X_t is a gambler's fortune after t tosses of a "fair" coin, where the gambler wins \$1 if the coin comes up heads and loses \$1 if the coin comes up tails. The gambler's conditional expected fortune after the next trial, given the history, is equal to his present fortune, so this sequence is a martingale. This is also known as D'Alembert system.

(4.) **Long Memory Process**

ARFIMA(n, ζ, s)

$$\Psi(L)(1-L)^{\zeta}(y_t - \mu_t) = \Theta(L)\varepsilon_t,$$

where the operator $(1-L)^{\zeta}$ accounts for the long memory of the process and is defined as:

$$(1-L)^{\zeta} = \sum_{k=0}^{\infty} \frac{\Gamma(k-\zeta)}{\Gamma(k+1)\Gamma(-\zeta)} L^k$$

with $-0.5 < \zeta < 0.5$. $\Gamma(.)$ denoting the Gamma function. The truncation order of the infinite summation is set to $t-1$.

(1.)

GPH estimation procedure-this was suggested by Geweke and Porter-Hudak (1983). Based on a regression of the ordinates, $w_1, w_2, \cdots, w_m$ of the log spectral density on trigonometric function as follows,

$$\ln_n \{(w_j)\} = a - 2d \ln \left(2 sin \frac{w_j}{2}\right) + u_j \quad j = 1, 2, \cdots, n^{1/2}$$

which holds with u_j approximately i.i.d. for low-order Fourier frequencies $w_j = 2\pi_j/n$, and $I_n(.)$ denotes the periodogram of the series, GPH suggests a semiparametric estimator of the fractional differencing estimator, d. Then, use time-domain maximum likelihood to the ARMA coefficients. In general, GPH is widely used in empirical works, however, Agiakoglou *et al.* (1993) point out that this two step periodogram regression method may cause large finite sample biases as short memory ARMA components are also present.

(2.) Maximum likelihood estimation (MLE)-The most efficient estimation procedure for fractionally differenced models is presumably , hence, MLE has been considered in several literature to estimate the joint estimation of the parameters in the ARFIMA (p, d, q) model (1) under the assumption of normality. The (p+q+3)-dimemsional vector of parameters is $\lambda' = (\mu, \beta')$, where $\beta' = (d, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma^2)$. Sowell (1992) derives the unconditional exact likelihood function for a normally distributed stationary fractionally integrated time series and uses recursive procedures allowing for evaluation of likelihood function. The log-likelihood is

$$\zeta = -(\frac{T}{2})log(2\pi) - (1/2)log(\Omega) - (1/2)Y'\Omega^{-1}Y,$$

where $\Omega_{ij} = \gamma_{|i-j|}$ and Y represents a T-dimensional vector of the observations on the process y_t. However, Sowell's is computationally demanding and for the theoretical mean parameter μ to be either zero or known. For the case of μ unkown, the formal proof of asymptotic normality and the appropriate rates of convergence of the MLE for the ARFIMA (p, d, q) is due to Dahlhaus (1988, 1989) for the $0 < d < 0.5$ case. As an alternative, no matter the mean parameter μ is known or unknown, on the framework of the autoregressive representation defined as

$$\sum_{j=0}^{\infty} a_j X_{t-j} = \epsilon,$$

Beran (1995) provides a simple algorithm of maximum likehood method for calculating the estimate of d and the ARMA parameters as the following procedures: with respect to η, minimize the sum of the squared residuals

$$S = \sum_{t=2}^{T} e_t^2(\eta)$$

where $e_t(\eta)$ is defined by equation (9) or (12) in Beran (1995), and $\eta = (\phi_1, \phi_2, \cdots, \phi_p, \phi_1, \phi_2, \phi_q)$ are the corresponding coefficients of the ARMA process.

ARCH Framework

y_t $(t = 1, \ldots, n)$, is typically modelled as follows:

$$y_t = m_t(\eta) + \varepsilon_t$$

$$\varepsilon_t = \sigma_t(\eta) z_t$$

$$m_t(\eta) = c(\eta|\Omega_{t-1})$$

$$\sigma_t(\eta) = h(\eta|\Omega_{t-1}).$$

where $c(.|\Omega_{t-1})$ and $h(.|\Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t-1$, i.e. $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$), depending on unknown vectors of parameters η, and z_t is an *i.i.d.* process with $\mathrm{E(z_t)} = 0$ and $\mathrm{Var(z_t)} = 1$.

Moments of ARCH processes

(1.) Let us denote by $\mathrm{E_{t-1}}(.)$ the conditional expectation at time $t-1$ and $V_{t-1}(.)$ the conditional variance at time $t-1$.

(2.) By definition, z_t is such that $\mathrm{E_{t-1}(z_t)} = 0$, $\mathrm{E_{t-1}(z_t^2)} = 1$.

(3.) Because the error term (ε_t) of the return equation is defined as the product of σ_t and z_t, it follows that it is uncorrelated but not independent as there is autocorrelation in the squares of ε_t.

Moments of ARCH processes

(1). $\mathrm{E_{t-1}(y_t) = m_t}(\eta)$ and $V_{t-1}(y_t) = \sigma_t^2(\eta)$.

(2). Indeed, $\mathrm{E_{t-1}(y_t) = E_{t-1}(m_t}(\eta)) + \mathrm{E_{t-1}(z_t}\sigma_\mathrm{t}(\eta)) = \mathrm{m_t}(\eta)$.

(3) For the conditional variance, $V_{t-1}(y_t) = V_{t-1}(\varepsilon_t) = \mathrm{E_{t-1}}(\varepsilon_\mathrm{t}^2) = \mathrm{E_{t-1}(z_t^2}\sigma_\mathrm{t}^2(\eta)) = 1 \times \sigma_\mathrm{t}^2(\eta) = \sigma_\mathrm{t}^2(\eta)$ because $\sigma_t^2(\eta)$ is entirely conditioned by the information known at time $t-1$.

(4). $\Rightarrow \sigma_t^2(\eta)$ is the conditional variance of the return process.

(5). The main feature of the ARCH model is to introduce a direct relationship between the conditional variance and the past (squared) returns

ARCH-in-Mean

Financial theory tells us that certain sources of risk are priced by the market.

(1). Assets with more 'risk' may provide higher average returns to compensate it.

(2). If σ_t^2 is an appropriate measure of risk, the conditional variance may enter the conditional mean function of y_t.

(3). ARCH-M of Engle, Lilien and Roberts (1987):

$$y_t = m_t + \kappa\sigma_t^2 + \varepsilon_t,$$

where σ_t^2 follows an ARCH-type model.

(4). Alternatively: $y_t = m_t + \kappa\sigma_t + \varepsilon_t$.

ARCH effects

(1.) We have seen that a suitable model for asset returns should include an uncorrelated error term, but which could possibly feature dependence in the squares, absolute values, or other functions of the returns.

(2.) The ARCH model aims to address this issue by introducing temporal dependence between the squares of the error term.