Econ 512: Financial Econometrics
Time Series Concepts

Eric Zivot

March 30, 2009

1 Reading

- MFTS, chapter 3
- AFTS, chapter 2
- APDVP, chapter 3 and 4 (sections 9 - 14)

2 Univariate Time Series

First, we present concepts useful for the analysis of univariate time series

2.1 Stationary and Ergodic Time Series

Let $\{y_t\} = \{\ldots y_{t-1}, y_t, y_{t+1}, \ldots\}$ denote a sequence of random variables indexed by some time subscript t. Call such a sequence of random variables a *time series*.

The time series $\{y_t\}$ is *covariance stationary* if

$$\begin{aligned} E[y_t] &= \mu \text{ for all } t \\ cov(y_t, y_{t-j}) &= E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \text{ for all } t \text{ and any } j \end{aligned}$$

The parameter γ_j is called the j^{th} order or lag j *autocovariance* of $\{y_t\}$ and a plot of γ_j against j is called the *autocovariance function*.

The *autocorrelations* of $\{y_t\}$ are defined by

$$\rho_j = \frac{cov(y_t, y_{t-j})}{\sqrt{var(y_t)var(y_{t-j})}} = \frac{\gamma_j}{\gamma_0}$$

and a plot of ρ_j against j is called the *autocorrelation function* (ACF)

The lag j *sample autocovariance* and lag j *sample autocorrelation* are defined as

$$\begin{aligned}\hat{\gamma}_j &= \frac{1}{T}\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y}) \\ \hat{\rho}_j &= \frac{\hat{\gamma}_j}{\hat{\gamma}_0}\end{aligned}$$

where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ is the sample mean.

The sample ACF (SACF) is a plot of $\hat{\rho}_j$ against j.

A stationary time series $\{y_t\}$ is *ergodic* if sample moments converge in probability to population moments; i.e. if $\bar{y} \xrightarrow{p} \mu, \hat{\gamma}_j \xrightarrow{p} \gamma_j$ and $\hat{\rho}_j \xrightarrow{p} \rho_j$.

Example: White noise (GWN) processes

Perhaps the most simple stationary time series is the *independent Gaussian white noise* process $y_t \sim iid\ N(0, \sigma^2) \equiv GWN(0, \sigma^2)$. This process has $\mu = \gamma_j = \rho_j = 0$ $(j \neq 0)$.

Two slightly more general processes are the independent *white noise* (IWN) process, $y_t \sim IWN(0, \sigma^2)$, and the *white noise* (WN) process, $y_t \sim WN(0, \sigma^2)$.

Both processes have mean zero and variance σ^2, but the IWN process has independent increments, whereas the WN process has uncorrelated increments.

The SACF is typically shown with 95% confidence limits about zero. These limits are based on the result that if $\{y_t\} \sim iid\,(0, \sigma^2)$ then

$$\hat{\rho}_j \overset{A}{\sim} N\left(0, \frac{1}{T}\right),\ j > 0.$$

The notation $\hat{\rho}_j \overset{A}{\sim} N\left(0, \frac{1}{T}\right)$ means that the distribution of $\hat{\rho}_j$ is approximated by normal distribution with mean 0 and variance $\frac{1}{T}$ and is based on the central limit theorem result $\sqrt{T}\hat{\rho}_j \xrightarrow{d} N\,(0, 1)$. The 95% limits about zero are then $\pm\frac{1.96}{\sqrt{T}}$.

2.1.1 Testing for White Noise

Consider testing the null hypothesis

$$H_0 : y_t \sim WN(0, \sigma^2)$$

Under the null, all of the autocorrelations ρ_j for $j > 0$ are zero. To test this null, Box and Pierce (1970) suggested the *Q-statistic*

$$Q(k) = T \sum_{j=1}^{k} \hat{\rho}_j^2$$

Under the null, $Q(k)$ is asymptotically distributed $\chi^2(k)$. In a finite sample, the Q-statistic may not be well approximated by the $\chi^2(k)$. Ljung and Box (1978) suggested the *modified Q-statistic*

$$MQ(k) = T(T+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j}$$

2.2 Linear Processes and ARMA Models

Wold's decomposition theorem (c.f. Fuller (1996) pg. 96) states that any covariance stationary time series $\{y_t\}$ has a *linear process* or infinite order moving average representation of the form

$$
\begin{aligned}
y_t &= \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \ \psi_0 = 1, \ \sum_{k=0}^{\infty} \psi_k^2 < \infty \\
\varepsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

In the Wold form, it can be shown that

$$
\begin{aligned}
E[y_t] &= \mu,\ \gamma_0 = var(y_t) = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2 \\
\gamma_j &= cov(y_t, y_{t-j}) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j} \\
\rho_j &= \frac{\sum_{k=0}^{\infty} \psi_k \psi_{k+j}}{\sum_{k=0}^{\infty} \psi_k^2}
\end{aligned}
$$

The moving average weights in the Wold form are also called *impulse responses* since

$$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s, s = 1, 2, \ldots$$

For a stationary and ergodic time series $\lim_{s\to\infty} \psi_s = 0$ and the *long-run cumulative impulse response* $\sum_{s=0}^{\infty} \psi_s < \infty$.

A plot of ψ_s against s is called the *impulse response function* (IRF)

The general Wold form of a stationary and ergodic time series is handy for theoretical analysis but is not practically useful for estimation purposes. A very rich and practically useful class of stationary and ergodic processes is the *autoregressive-moving average* (ARMA) class of models made popular by Box and Jenkins (1976).

ARMA(p, q) models take the form of a pth order stochastic difference equation

$$\begin{aligned} y_t - \mu &= \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) \\ &\quad +\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

2.2.1 Lag Operator Notation

The lag operator L is defined such that for any time series $\{y_t\}$, $Ly_t = y_{t-1}$. It has the following properties: $L^2 y_t = L \cdot Ly_t = y_{t-2}$, $L^0 = 1$ and $L^{-1} y_t = y_{t+1}$. The operator $\Delta = 1 - L$ creates the first difference of a time series: $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$. The ARMA$(p, q)$ model may be compactly expressed using lag polynomials. Define $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$. Then the ARMA model may be expressed as

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

Similarly, the Wold representation in lag operator notation is

$$\begin{aligned} y_t &= \mu + \psi(L)\varepsilon_t, \ \psi(L) = \sum_{k=0}^{\infty} \psi_k L^k, \ \psi_0 = 1 \\ \psi(1) &= \sum_{k=0}^{\infty} \psi_k \end{aligned}$$

2.3 Autoregressive Models

2.3.1 AR(1) Model

A commonly used stationary and ergodic time series in financial modeling is the AR(1) process

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t,\ t = 1, \ldots, T$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$. The above representation is called the *mean-adjusted form*. The *characteristic equation* for the AR(1) is

$$\phi(z) = 1 - \phi z = 0$$

so that the root is $z = \frac{1}{\phi}$.

Stationarity is satisfied provided the absolute value of the root of the characteristic equation is greater than one: $|\frac{1}{\phi}| > 1$ or $|\phi| < 1$. In this case, it is easy to show that $E[y_t] = \mu$, $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$, $\psi_j = \rho_j = \phi^j$ and the Wold representation is

$$y_t = \mu + \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}.$$

In a stationary AR(1) model, $\{y_t\}$ exhibits *mean-reverting* behavior. That is, $\{y_t\}$ fluctuates about the mean value $\mu = 1$. The ACF and IRF decay at a geometric rate.

The decay rate of the IRF is sometimes reported as a *half-life* – the lag j^{half} at which the IRF reaches $\frac{1}{2}$. For the AR(1) with positive ϕ, it can be shown that

$$j^{half} = \ln(0.5)/\ln(\phi)$$

2.3.2 AR(p) Models

The AR(p) model in mean-adjusted form is

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$$

or, in lag operator notation,

$$\phi(L)(y_t - \mu) = \varepsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$. The autoregressive form is

$$\phi(L)y_t = c + \varepsilon_t.$$

It can be shown that the AR(p) is stationary and ergodic provided the roots of the *characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

lie outside the complex unit circle (have modulus greater than one). A necessary condition for stationarity that is useful in practice is that $|\phi_1 + \cdots + \phi_p| < 1$. If $\phi(z) = 0$ has complex roots then y_t will exhibit sinusoidal behavior. In the stationary AR(p), the constant in the autoregressive form is equal to $\mu(1 - \phi_1 - \cdots - \phi_p)$.

2.3.3 Partial Autocorrelation Function

The *partial autocorrelation function* (PACF) is a useful tool to help identify AR(p) models. The PACF is based on estimating the sequence of AR models

$$
\begin{aligned}
z_t &= \phi_{11} z_{t-1} + \varepsilon_{1t} \\
z_t &= \phi_{21} z_{t-1} + \phi_{22} z_{t-2} + \varepsilon_{2t} \\
&\vdots \\
z_t &= \phi_{p1} z_{t-1} + \phi_{p2} z_{t-2} + \cdots + \phi_{pp} z_{t-p} + \varepsilon_{pt}
\end{aligned}
$$

where $z_t = y_t - \mu$ is the demeaned data. The coefficients ϕ_{jj} for $j = 1, \ldots, p$ (i.e., the last coefficients in each AR(p) model) are called the *partial autocorrelation coefficients*.

Note: for an AR(p) model, all of the first p partial autocorrelation coefficients are non-zero, and the rest are zero for $j > p$.

2.4 Moving Average Models

2.4.1 MA(1) Model

The MA(1) model has the form

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim WN(0, \sigma^2)$$

For any finite θ the MA(1) is stationary and ergodic. The moments are $E[y_t] = \mu$, $\gamma_0 = \sigma^2(1+\theta^2)$, $\gamma_1 = \sigma^2\theta$, $\gamma_j = 0$ for $j > 1$ and $\rho_1 = \theta/(1+\theta^2)$. Hence, the ACF of an MA(1) process cuts off at lag one, and the maximum value of this correlation is ± 0.5.

There is an identification problem with the MA(1) model since $\theta = 1/\theta$ produce the same value of ρ_1. The MA(1) is called *invertible* if $|\theta| < 1$ and is called *non-invertible* if $|\theta| \geq 1$. In the invertible MA(1), the error term ε_t has an infinite order AR representation of the form

$$\varepsilon_t = \sum_{j=0}^{\infty} \theta^{*j}(y_{t-j} - \mu)$$

where $\theta^* = -\theta$ so that ε_t may be thought of as a prediction error based on past values of y_t.

MA(1) models often arise through data transformations like aggregation and differencing. For example, consider the signal plus noise model

$$
\begin{aligned}
y_t &= z_t + \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \\
z_t &= z_{t-1} + \eta_t,\ \eta_t \sim WN(0, \sigma_\eta^2)
\end{aligned}
$$

where ε_t and η_t are independent. For example, z_t could represent the fundamental value of an asset price and ε_t could represent an iid deviation about the fundamental price. A stationary representation requires differencing y_t:

$$\Delta y_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}$$

It can be shown that Δy_t is an MA(1) process with $\theta = \frac{-(q+2)+\sqrt{q^2+4q}}{2}$ where $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ is the signal-to-noise ratio and $\rho_1 = \frac{-1}{q+2} < 0$.

2.4.2 MA(q) Model

The MA(q) model has the form

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \text{ where } \varepsilon_t \sim WN(0, \sigma^2)$$

The MA(q) model is stationary and ergodic provided $\theta_1, \ldots, \theta_q$ are finite. It is *invertible* if all of the roots of the MA characteristic polynomial

$$\theta(z) = 1 + \theta_1 z + \cdots \theta_q z^q = 0$$

lie outside the complex unit circle. The moments of the MA(q) are

$$\begin{aligned} E[y_t] &= \mu \\ \gamma_0 &= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \\ \gamma_j &= \begin{cases} \left(\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q \theta_{q-j}\right)\sigma^2 \text{ for } j = 1, 2, \ldots, q \\ 0 \text{ for } j > q \end{cases} \end{aligned}$$

Hence, the ACF of an MA(q) is non-zero up to lag q and is zero afterwards.

Example: Overlapping returns and MA(q) models

MA(q) models often arise in finance through data aggregation transformations. For example, let $R_t = \ln(P_t/P_{t-1})$ denote the monthly continuously compounded return on an asset with price P_t. Define the annual return at time t using monthly returns as $R_t(12) = \ln(P_t/P_{t-12}) = \sum_{j=0}^{11} R_{t-j}$. Suppose $R_t \sim WN(\mu, \sigma^2)$ and consider a sample of monthly returns of size T, $\{R_1, R_2, \ldots, R_T\}$.

A sample of annual returns may be created using *overlapping* or *non-overlapping* returns. Let $\{R_{12}(12), R_{13}(12), \ldots, R_T(12)\}$ denote a sample of $T^* = T - 11$ monthly overlapping annual returns and $\{R_{12}(12), R_{24}(12), \ldots, R_T(12)\}$ denote a sample of $T/12$ non-overlapping annual returns.

Researchers often use overlapping returns in analysis due to the apparent larger sample size. One must be careful using overlapping returns because the monthly annual return sequence $\{R_t(12)\}$ is not a white noise process even if the monthly return sequence $\{R_t\}$ is. To see this, straightforward calculations give

$$
\begin{aligned}
E[R_t(12)] &= 12\mu \\
\gamma_0 &= var(R_t(12)) = 12\sigma^2 \\
\gamma_j &= cov(R_t(12), R_{t-j}(12)) = (12 - j)\sigma^2 \text{ for } j < 12 \\
\gamma_j &= 0 \text{ for } j \geq 12
\end{aligned}
$$

Since $\gamma_j = 0$ for $j \geq 12$ notice that $\{R_t(12)\}$ behaves like an MA(11) process

$$
\begin{aligned}
R_t(12) &= 12\mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_{11}\varepsilon_{t-11} \\
\varepsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

2.5 ARMA(p,q) Models

The regression formulation of the ARMA(p,q) model is

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta\varepsilon_{t-1} + \cdots + \theta\varepsilon_{t-q}$$

It is stationary and ergodic if the roots of the characteristic equation $\phi(z) = 0$ lie outside the complex unit circle, and it is invertible if the roots of the MA characteristic polynomial $\theta(z) = 0$ lie outside the unit circle.

It is assumed that the polynomials $\phi(z) = 0$ and $\theta(z) = 0$ do not have canceling or common factors.

A stationary and ergodic ARMA(p, q) process has a mean equal to

$$\mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p}$$

and its autocovariances, autocorrelations and impulse response weights satisfy the recursive relationships

$$\begin{aligned} \gamma_j &= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p} \\ \rho_j &= \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p} \\ \psi_j &= \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \cdots + \phi_p \psi_{j-p} \end{aligned}$$

ARMA(p, q) models often arise from certain aggregation transformations of simple time series models. An important result due to Granger and Morris (1976) is that if y_{1t} is an ARMA(p_1, q_1) process and y_{2t} is an ARMA(p_2, q_2) process, which may be contemporaneously correlated with y_{1t}, then $y_{1t} + y_{2t}$ is an ARMA(p, q) process with $p = p_1 + p_2$ and $q = \max(p_1 + q_2, q_1 + p_2)$.

For example, if y_{1t} is an AR(1) process and y_2 is a AR(1) process, then $y_1 + y_2$ is an ARMA(2,1) process.

High order ARMA(p, q) processes are difficult to identify and estimate in practice and are rarely used in the analysis of financial data. Low order ARMA(p, q) models with p and q less than three are generally sufficient for the analysis of financial data.

2.5.1 ARIMA(p, d, q) Models

The specification of the ARMA(p, q) model assumes that y_t is stationary and ergodic. If y_t is a trending variable like an asset price or a macroeconomic aggregate like real GDP, then y_t must be transformed to stationary form by eliminating the trend. Box and Jenkins (1976) advocate removal of trends by differencing.

Let $\Delta = 1 - L$ denote the *difference operator*. If there is a linear trend in y_t then the first difference $\Delta y_t = y_t - y_{t-1}$ will not have a trend. If there is a quadratic trend in y_t, then Δy_t will contain a linear trend but the second difference $\Delta^2 y_t = (1-2L+L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$ will not have a trend.

The class of ARMA(p, q) models where the trends have been transformed by differencing d times is denoted ARIMA(p, d, q)

2.6 Estimation of ARMA Models and Forecasting

ARMA(p, q) models are generally estimated using the technique of maximum likelihood, which is usually accomplished by putting the ARMA(p, q) in state-space form from which the prediction error decomposition of the log-likelihood function may be constructed.

The *exact likelihood* utilizes the stationary distribution of the initial values in the construction of the likelihood. The *conditional likelihood* treats the p initial values of y_t as fixed and often sets the q initial values of ε_t to zero. The exact maximum likelihood estimates (MLEs) maximize the exact log-likelihood, and the conditional MLEs maximize the conditional log-likelihood.

The exact and conditional MLEs are asymptotically equivalent but can differ substantially in small samples, especially for models that are close to being nonstationary or noninvertible.

2.6.1 Model Selection Criteria

Before an ARMA(p,q) may be estimated for a time series y_t, the AR and MA orders p and q must be determined by visually inspecting the SACF and SPACF for y_t. Alternatively, statistical *model selection criteria* may be used. The idea is to fit all ARMA(p,q) models with orders $p \leq p_{\text{max}}$ and $q \leq q_{\text{max}}$ and choose the values of p and q which minimizes some model selection criteria. Model selection criteria for ARMA(p,q) models have the form

$$MSC(p,q) = \ln(\tilde{\sigma}^2(p,q)) + c_T \cdot \varphi(p,q)$$

where $\tilde{\sigma}^2(p,q)$ is the MLE of $var(\varepsilon_t) = \sigma^2$ *without a degrees of freedom correction* from the ARMA(p,q) model, c_T is a sequence indexed by the sample size T, and $\varphi(p,q)$ is a penalty function which penalizes large ARMA(p,q) models.

The two most common information criteria are the Akaike (AIC) and Schwarz-Bayesian (BIC):

$$\begin{aligned} AIC(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{2}{T}(p+q) \\ BIC(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{\ln T}{T}(p+q) \end{aligned}$$

The AIC criterion asymptotically overestimates the order with positive probability, whereas the BIC estimate the order consistently under fairly general conditions if the true orders p and q are less than or equal to $p_{\max}$ and $q_{\max}$. However, in finite samples the BIC generally shares no particular advantage over the AIC.

2.6.2 Forecasting Algorithm

Forecasts from an ARIMA(p, d, q) model are straightforward. The model is put in state space form, and optimal h-step ahead forecasts along with forecast standard errors (not adjusted for parameter uncertainty) are produced using the Kalman filter algorithm. Details of the method are given in Harvey (1993).

2.7 Martingales and Martingale Difference Sequences

Let $\{y_t\}$ denote a sequence of random variables and let $I_t = \{y_t, y_{t-1}, \ldots\}$ denote a set of conditioning information or *information set* based on the past history of y_t. The sequence $\{y_t, I_t\}$ is called a *martingale* if

- $I_{t-1} \subset I_t$ (I_t is a filtration)
- $E[|y_t|] < \infty$
- $E[y_t | I_{t-1}] = y_{t-1}$ (martingale property)

The most common example of a martingale is the random walk model

$$y_t = y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim WN(0, \sigma^2)$$

where y_0 is a fixed initial value. Letting $I_t = \{y_t, \ldots, y_0\}$ implies $E[y_t|I_{t-1}] = y_{t-1}$ since $E[\varepsilon_t|I_{t-1}] = 0$.

Law of Iterated Expectations. Let $\{Y_t, I_t\}$ be a martingale. Then

$$\begin{aligned} E[Y_t|I_{t-2}] &= E[E[Y_t|I_{t-1}]|I_{t-2}] \\ &= E[Y_{t-1}|I_{t-2}] = Y_{t-2} \end{aligned}$$

It follows that

$$E[Y_t|I_{t-k}] = Y_{t-k}$$

In general, for information sets I_t and J_t such that $I_t \subset J_t$ (J_t is the bigger info set). The *Law of Iterated Expectations* says

$$E[Y|I_t] = E[E[X|J_t]|I_t]$$

Let $\{\varepsilon_t\}$ be a sequence of random variables with an associated information set I_t. The sequence $\{\varepsilon_t, I_t\}$ is called a *martingale difference sequence* (MDS) if

- $I_{t-1} \subset I_t$
- $E[\varepsilon_t | I_{t-1}] = 0$ (MDS property)

If $\{y_t, I_t\}$ is a martingale, a MDS $\{\varepsilon_t, I_t\}$ may be constructed by defining

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

By construction, a MDS is an uncorrelated process. This follows from the *law of iterated expectations.* To see this, for any $k > 0$

$$\begin{aligned} E[\varepsilon_t \varepsilon_{t-k}] &= E[E[\varepsilon_t \varepsilon_{t-k} | I_{t-1}]] \\ &= E[\varepsilon_{t-k} E[\varepsilon_t | I_{t-1}]] \\ &= 0 \end{aligned}$$

In fact, if z_n is any function of the past history of ε_t so that $z_n \in I_{t-1}$ then

$$E[\varepsilon_t z_n] = 0$$

Example: ARCH process

A well known stylized fact about high frequency financial asset returns is that volatility appears to be autocorrelated. A simple model to capture such volatility autocorrelation is the ARCH process due to Engle (1982). To illustrate, let r_t denote the daily return on an asset and assume that $E[r_t] = 0$. An ARCH(1) model for r_t is

$$\begin{aligned} r_t &= \sigma_t z_t \\ z_t &\sim iid\ N(0,1) \\ \sigma_t^2 &= \omega + \alpha r_{t-1}^2 \end{aligned}$$

where $\omega > 0$ and $0 < \alpha < 1$. Let $I_t = \{r_t, \ldots\}$.

To see that $\{r_t, I_t\}$ is a MDS, note that

$$\begin{aligned} E[r_t|I_{t-1}] &= E[z_t\sigma_t|I_{t-1}] \\ &= \sigma_t E[z_t|I_{t-1}] \\ &= 0 \end{aligned}$$

Since r_t is a MDS, it is an uncorrelated process. Provided $|\alpha| < 1$, r_t is a mean zero covariance stationary process. The unconditional variance of r_t is given by

$$\begin{aligned} var(r_t) &= E[r_t^2] = E[E[z_t^2 \sigma_t^2 | I_{t-1}]] \\ &= E[\sigma_t^2 E[z_t^2 | I_{t-1}] = E[\sigma_t^2] \end{aligned}$$

since $E[z_t^2 | I_{t-1}] = 1$.

$E[\sigma_t^2]$ may be expressed as

$$E[\sigma_t^2] = \frac{\omega}{1-\alpha}$$

Furthermore, by adding r_t^2 to both sides of $\sigma_t^2 = \omega + \alpha r_{t-1}^2$ and rearranging it follows that r_t^2 has an AR(1) representation of the form

$$r_t^2 = \omega + \alpha r_{t-1}^2 + v_t$$

where $v_t = r_t^2 - \sigma_t^2$ is a MDS.

2.8 Long-run Variance

Let y_t be a stationary and ergodic time series. Anderson's central limit theorem for stationary and ergodic processes (c.f. Hamilton (1994) pg. 195) states

$$\sqrt{T}(\bar{y} - \mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty} \gamma_j)$$

or

$$\bar{y} \overset{A}{\sim} N\left(\mu, \frac{1}{T} \sum_{j=-\infty}^{\infty} \gamma_j\right)$$

The sample size, T, times the *asymptotic variance* of the sample mean is often called the *long-run variance* of y_t :

$$lrv(y_t) = T \cdot avar(\bar{y}) = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j.$$

2.8.1 Estimating the Long-Run Variance

If y_t is a linear process, it may be shown that

$$\sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2 = \sigma^2 \psi(1)^2$$

and so

$$lrv(y_t) = \sigma^2 \psi(1)^2$$

Further, if $y_t \sim \text{ARMA}(p, q)$ then

$$\psi(1) = \frac{1 + \theta_1 + \cdots + \theta_q}{1 - \phi_1 - \cdots - \phi_p} = \frac{\theta(1)}{\phi(1)}$$

so that

$$lrv(y_t) = \frac{\sigma^2 \theta(1)^2}{\phi(1)^2}.$$

Alternatively, the ARMA(p, q) process may be approximated by a high order AR(p^*) process

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_{p^*} y_{t-p^*} + \varepsilon_t$$

where the lag length p^* is chosen such that ε_t is uncorrelated. This gives rise to the *autoregressive long-run variance* estimate

$$lrv_{AR}(y_t) = \frac{\sigma^2}{\phi^*(1)^2}.$$

A consistent estimate of $lrv(y_t)$ may also be computed using some non-parametric methods. An estimator made popular by Newey and West (1987) is the weighted autocovariance estimator

$$\widehat{lrv}_{NW}(y_t) = \hat{\gamma}_0 + 2 \sum_{j=1}^{M_T} w_{j,T} \cdot \hat{\gamma}_j$$

where $w_{j,T}$ are weights which sum to unity and M_T is a truncation lag parameter that satisfies $M_T = O(T^{1/3})$.

3 Univariate Nonstationary Time Series

A univariate time series process $\{y_t\}$ is called *nonstationary* if it is not stationary. Since a stationary process has time invariant moments, a nonstationary process must have some time dependent moments. The most common forms of nonstationarity are caused by time dependence in the mean and variance.

3.0.2 Trend Stationary Process

$\{y_t\}$ is a *trend stationary* process if it has the form

$$y_t = TD_t + x_t$$

where TD_t are deterministic trend terms (constant, trend, seasonal dummies etc) that depend on t and $\{x_t\}$ is stationary. The series y_t is nonstationary because $E[TD_t] = TD_t$ which depends on t. Since x_t is stationary, y_t never deviates too far away from the deterministic trend TD_t. Hence, y_t exhibits *trend reversion*. If TD_t were known, y_t may be transformed to a stationary process by subtracting off the deterministic trend terms:

$$x_t = y_t - TD_t$$

Example: Trend stationary AR(1)

A trend stationary AR(1) process with $TD_t = \mu + \delta t$ may be expressed in three equivalent ways

$$\begin{aligned} y_t &= \mu + \delta t + u_t, u_t = \phi u_{t-1} + \varepsilon_t \\ y_t - \mu - \delta t &= \phi(y_{t-1} - \mu - \delta(t-1)) + \varepsilon_t \\ y_t &= c + \beta t + \phi y_{t-1} + \varepsilon_t \end{aligned}$$

where $|\phi| < 1$, $c = \mu(1-\phi) + \delta$, $\beta = \delta(1-\phi)t$ and $\varepsilon_t \sim WN(0, \sigma^2)$.

3.0.3 Integrated Processes

$\{y_t\}$ is an *integrated process* of order 1, denoted $y_t \sim I(1)$, if it has the form

$$y_t = y_{t-1} + u_t$$

where u_t is a stationary time series. Clearly, the first difference of y_t is stationary

$$\Delta y_t = u_t$$

Because of the above property, $I(1)$ processes are sometimes called *difference stationary* processes.

Starting at y_0, by recursive substitution y_t has the representation of an *integrated sum* of stationary innovations

$$y_t = y_0 + \sum_{j=1}^{t} u_j.$$

The integrated sum $\sum_{j=1}^{t} u_j$ is called a *stochastic trend* and is denoted TS_t. Notice that

$$TS_t = TS_{t-1} + u_t,\ TS_0 = 0$$

If $u_t \sim IWN(0, \sigma^2)$ in (**??**) then y_t is called a *random walk*. In general, an $I(1)$ process can have serially correlated and heteroskedastic innovations u_t.

If y_t is a random walk and assuming y_0 is fixed then it can be shown that

$$\begin{aligned}
\gamma_0 &= \sigma^2 t \\
\gamma_j &= (t-j)\sigma^2 \\
\rho_j &= \sqrt{\frac{t-j}{t}}
\end{aligned}$$

which clearly shows that y_t is nonstationary. Also, if t is large relative to j then $\rho_j \approx 1$. Hence, for an $I(1)$ process, the ACF does not decay at a geometric rate but at a linear rate as j increases.

An $I(1)$ process with drift has the form

$$y_t = \mu + y_{t-1} + u_t, \text{ where } u_t \sim I(0)$$

Starting at $t = 0$ an $I(1)$ process with drift μ may be expressed as

$$\begin{aligned} y_t &= y_0 + \mu t + \sum_{j=1}^{t} u_t \\ &= TD_t + TS_t \end{aligned}$$

so that it may be thought of as being composed of a deterministic linear trend $TD_t = y_0 + \mu t$ as well as a stochastic trend $TS_t = \sum_{j=1}^{t} u_j$.

An $I(d)$ process $\{y_t\}$ is one in which $\Delta^d y_t \sim I(0)$.

In finance and economics data series are rarely modeled as $I(d)$ process with $d > 2$.

Just as an $I(1)$ process with drift contains a linear deterministic trend, an $I(2)$ process with drift will contain a quadratic trend.

4 Long Memory Time Series

If a time series y_t is $I(0)$ then its ACF declines at a geometric rate. As a result, $I(0)$ process have *short memory* since observations far apart in time are essentially independent.

Conversely, if y_t is $I(1)$ then its ACF declines at a linear rate and observations far apart in time are not independent.

In between $I(0)$ and $I(1)$ processes are so-called *fractionally integrated* $I(d)$ process where $0 < d < 1$. The ACF for a fractionally integrated processes declines at a polynomial (hyperbolic) rate, which implies that observations far apart in time may exhibit weak but non-zero correlation. This weak correlation between observations far apart is often referred to as *long memory*.

A fractionally integrated white noise process y_t has the form

$$(1-L)^d y_t = \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma^2)$$

where $(1-L)^d$ has the binomial series expansion representation (valid for any $d > -1$)

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \\ &= 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots\end{aligned}$$

If $d = 1$ then y_t is a random walk and if $d = 0$ then y_t is white noise. For $0 < d < 1$ it can be shown that

$$\rho_k \propto k^{2d-1}$$

as $k \to \infty$ so that the ACF for y_t declines hyperbolically to zero at a speed that depends on d. Further, it can be shown y_t is stationary and ergodic for $0 < d < 0.5$ and that the variance of y_t is infinite for $0.5 \leq d < 1$.

A fractionally integrated process with stationary and ergodic ARMA(p, q) errors

$$(1 - L)^d y_t = u_t, \; u_t \sim ARMA(p, q)$$

is called an *autoregressive fractionally integrated moving average* (ARFIMA) process.

5 Statistical Tests for Long Memory

Given the scaling property of the autocorrelation function and the fractionally integrated process representation of a long memory time series, various tests have been proposed to determine the existence of long memory in a time series.

However, it is important to note that the definition of long memory does not dictate the general behavior of the autocorrelation function.

Instead, it only specifies the asymptotic behavior when $k \longrightarrow \infty$. What this means is that for a long memory process, it is not necessary for the autocorrelation to remain significant at large lags as long as the autocorrelation function decays slowly.

5.1 R/S Statistic

The best-known test for long memory or long range dependence is probably the *rescaled range*, or *range over standard deviation*, or simply R/S statistic, which was originally proposed by Hurst (1951), and later refined by Mandelbrot and his coauthors.

The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. Specifically, consider a time series y_t, for $t = 1, \cdots, T$. The R/S statistic is defined as:

$$Q_T = \frac{1}{s_T}\left[\max_{1\le k\le T}\sum_{j=1}^{k}(y_j - \bar{y}) - \min_{1\le k\le T}\sum_{j=1}^{k}(y_j - \bar{y})\right]$$

where $\bar{y} = 1/T\sum_{i=1}^{T} y_i$ and $s_T = \sqrt{1/T\sum_{i=1}^{T}(y_i - \bar{y})^2}$.

If y_t's are i.i.d. normal random variables, then

$$\frac{1}{\sqrt{T}} Q_T \Rightarrow V$$

where $\Rightarrow$ denotes weak convergence and V is the range of a Brownian bridge on the unit interval. Lo (1991) gives selected quantiles of V.

Lo (1991) pointed out that the R/S statistic is not robust to short range dependence. In particular, if y_t is autocorrelated (has short memory) then the limiting distribution of $Q_T/\sqrt{T}$ is V scaled by the square root of the long run variance of y_t. To allow for short range dependence in y_t, Lo (1991) modified the R/S statistic as follows:

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) \right]$$

where the sample standard deviation is replaced by the square root of the Newey-West estimate of the long run variance with bandwidth q.

Lo (1991) showed that if there is short memory but no long memory in y_t, $\tilde{Q}_T$ also converges to V, the range of a Brownian bridge.

5.2 Estimation of d using R/S Statistic

When there is no long memory in a stationary time series, the R/S statistic converges to a random variable at rate $T^{1/2}$.

However, when the stationary process y_t has long memory, Mandelbrot (1975) showed that the R/S statistic converges to a random variable at rate $T^{d+1/2}$ for $d > 0$.

Based on this result, the log-log plot of the R/S statistic versus the sample size used should scatter around a straight line with slope 1/2 for a short memory time series.

In contrast, for a long memory time series, the log-log plot should scatter around a straight line with slope equal to $d + 1/2$, provided the sample size is large enough.

To use the R/S statistic to estimate the long memory parameter d, first compute the R/S statistic using k_1 consecutive observations in the sample, where k_1 should be large enough.

Then increase the number of observations by a factor of f; that is, compute the R/S statistic using $k_i = fk_{i-1}$ consecutive observations for $i = 2, \cdots, s$. Note that to obtain the R/S statistic with k_i consecutive observations, one can actually divide the sample into $[T/k_i]$ blocks and obtain $[T/k_i]$ different values, where $[\cdot]$ denotes the integer part of a real number. Obviously, the larger k_i is, the smaller $[T/k_i]$ is.

A line fit of all those R/S statistics versus k_i, $i = 1, \cdots, s$, on the log-log scale yields an estimate of $d + 1/2$.

6 Multivariate Time Series

Consider n time series variables $\{y_{1t}\}, \ldots, \{y_{nt}\}$. A *multivariate time series* is the $(n \times 1)$ vector time series $\{\mathbf{Y}_t\}$ where the i^{th} row of $\{\mathbf{Y}_t\}$ is $\{y_{it}\}$. That is, for any time t, $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$.

Multivariate time series analysis is used when one wants to model and explain the interactions and co-movements among a group of time series variables.

In finance, multivariate time series analysis is used to model systems of asset returns, asset prices and exchange rates, the term structure of interest rates, asset returns/prices, and economic variables etc.

Many of the time series concepts described previously for univariate time series carry over to multivariate time series in a natural way. Additionally, there are some important time series concepts that are particular to multivariate time series

6.1 Stationary and Ergodic Multivariate Time Series

A multivariate time series $\mathbf{Y}_t$ is covariance stationary and ergodic if all of its component time series are stationary and ergodic. The mean of $\mathbf{Y}_t$ is defined as the $(n \times 1)$ vector

$$E[\mathbf{Y}_t] = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$$

where $\mu_i = E[y_{it}]$ for $i = 1, \ldots, n$. The variance/covariance matrix of $\mathbf{Y}_t$ is the $(n \times n)$ matrix

$$\begin{aligned} var(\mathbf{Y}_t) &= \Gamma_0 = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})'] \\ &= \begin{pmatrix} var(y_{1t}) & cov(y_{1t}, y_{2t}) & \cdots & cov(y_{1t}, y_{nt}) \\ cov(y_{2t}, y_{1t}) & var(y_{2t}) & \cdots & cov(y_{2t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_{nt}, y_{1t}) & cov(y_{nt}, y_{2t}) & \cdots & var(y_{nt}) \end{pmatrix} \end{aligned}$$

The matrix $\mathbf{\Gamma}_0$ has elements $\gamma_{ij}^0 = cov(y_{it}, y_{jt})$. The correlation matrix of Y_t is the $(n \times n)$ matrix

$$corr(\mathbf{Y}_t) = \mathbf{R}_0 = \mathbf{D}^{-1}\mathbf{\Gamma}_0\mathbf{D}^{-1}$$

where $\mathbf{D}$ is an $(n \times n)$ diagonal matrix with j^{th} diagonal element $(\gamma_{jj}^0)^{1/2} = SD(y_{jt})$.

6.1.1 Cross Covariance and Correlation Matrices

The autocovariances and autocorrelations of y_{jt} for $j = 1, \ldots, n$ are defined as

$$\begin{aligned} \gamma_{jj}^{k} &= cov(y_{jt}, y_{jt-k}), \\ \rho_{jj}^{k} &= corr(y_{jt}, y_{jt-k}) = \frac{\gamma_{jj}^{k}}{\gamma_{jj}^{0}} \end{aligned}$$

and these are symmetric in k: $\gamma_{jj}^{k} = \gamma_{jj}^{-k}$, $\rho_{jj}^{k} = \rho_{jj}^{-k}$.

The *cross lag covariances* and *cross lag correlations* between y_{it} and y_{jt} are defined as

$$\begin{aligned} \gamma_{ij}^{k} &= cov(y_{it}, y_{jt-k}), \\ \rho_{ij}^{k} &= corr(y_{jt}, y_{jt-k}) = \frac{\gamma_{ij}^{k}}{\sqrt{\gamma_{ii}^{0}\gamma_{jj}^{0}}} \end{aligned}$$

and they are not necessarily symmetric in k. In general,

$$\gamma_{ij}^{k} = cov(y_{it}, y_{jt-k}) \neq cov(y_{it}, y_{jt+k}) = cov(y_{jt}, y_{it-k}) = \gamma_{ij}^{-k}$$

If $\gamma_{ij}^{k} \neq 0$ for some $k > 0$ then y_{jt} is said to *lead* y_{it}. Similarly, if $\gamma_{ij}^{-k} \neq 0$ for some $k > 0$ then y_{it} is said to *lead* y_{jt}. It is possible that y_{it} leads y_{jt} and vice-versa. In this case, there is said to be *feedback* between the two series.

All of the lag k cross covariances and correlations are summarized in the $(n \times n)$ lag k cross covariance and lag k cross correlation matrices

$$\begin{aligned}
\mathbf{\Gamma}_k &= E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-k} - \boldsymbol{\mu})'] \\
&= \begin{pmatrix} cov(y_{1t}, y_{1t-k}) & cov(y_{1t}, y_{2t-k}) & \cdots & cov(y_{1t}, y_{nt-k}) \\ cov(y_{2t}, y_{1t-k}) & cov(y_{2t}, y_{2t-k}) & \cdots & cov(y_{2t}, y_{nt-k}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_{nt}, y_{1t-k}) & cov(y_{nt}, y_{2t-k}) & \cdots & cov(y_{nt}, y_{nt-k}) \end{pmatrix} \\
\mathbf{R}_k &= \mathbf{D}^{-1}\mathbf{\Gamma}_k\mathbf{D}^{-1}
\end{aligned}$$

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are not symmetric in k but it is easy to show that $\mathbf{\Gamma}_{-k} = \mathbf{\Gamma}_k'$ and $\mathbf{R}_{-k} = \mathbf{R}_k'$.

6.2 Multivariate Wold Representation

Any $(n \times 1)$ covariance stationary multivariate time series $\mathbf{Y}_t$ has a Wold or linear process representation of the form

$$\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\
&= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}
\end{aligned}$$

where $\boldsymbol{\Psi}_0 = \mathbf{I}_n$ and $\boldsymbol{\varepsilon}_t$ is a multivariate white noise process with mean zero and variance matrix $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Sigma}$. $\boldsymbol{\Psi}_k$ is an $(n \times n)$ matrix with (i, j)th element ψ_{ij}^k.

In lag operator notation, the Wold form is

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L) \boldsymbol{\varepsilon}_t, \ \boldsymbol{\Psi}(L) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k L^k$$

The moments of $\mathbf{Y}_t$ are given by

$$E[\mathbf{Y}_t] = \boldsymbol{\mu},\ var(\mathbf{Y}_t) = \sum_{k=0}^{\infty} \mathbf{\Psi}_k \mathbf{\Sigma} \mathbf{\Psi}_k'$$

6.2.1 VAR Models

The most popular multivariate time series model is the *vector autoregressive* (VAR) model. The VAR model is a multivariate extension of the univariate autoregressive model. For example, a bivariate VAR(1) model has the form

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

or

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim iid \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$$

In the equations for y_1 and y_2, the lagged values of both y_1 and y_2 are present.

The general VAR(p) model for $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ has the form

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \mathbf{\Pi}_2 \mathbf{Y}_{t-2} + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \ t = 1, \ldots, T$$

6.3 Long Run Variance

Let $\mathbf{Y}_t$ be an $(n \times 1)$ stationary and ergodic multivariate time series with $E[\mathbf{Y}_t] = \boldsymbol{\mu}$. Anderson's central limit theorem for stationary and ergodic process states

$$\sqrt{T}(\bar{\mathbf{Y}} - \boldsymbol{\mu}) \xrightarrow{d} N\left(\mathbf{0}, \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j\right), \; \bar{\mathbf{Y}} \overset{A}{\sim} N\left(\boldsymbol{\mu}, \frac{1}{T} \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j\right)$$

Hence, the *long-run variance* of $\mathbf{Y}_t$ is T times the asymptotic variance of $\bar{\mathbf{Y}}$:

$$lrv(\mathbf{Y}_t) = T \cdot avar(\bar{\mathbf{Y}}) = \sum_{j=-\infty}^{\infty} \Gamma_j$$

Since $\mathbf{\Gamma}_{-j} = \mathbf{\Gamma}'_j$, $lrv(\mathbf{Y}_t)$ may be alternatively expressed as

$$lrv(\mathbf{Y}_t) = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty} (\mathbf{\Gamma}_j + \mathbf{\Gamma}'_j)$$