Linear and non-linear time series analysis

Marco Bee

marco.bee@unitn.it
Department of Economics and Management
University of Trento

Reaction
Trento Summer School
R: Economics in Action

September 5-8, 2017

Outline

Prices and returns

- Let P_t be the price of a financial asset. The (one-period) net return is given by

$$R_{t|t-1} = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

- The gross return is

$$R^g_{t|t-1} = \frac{P_t}{P_{t-1}}.$$

- The logarithmic return is

$$r_{t|t-1} = \log(R^g_{t|t-1}) = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1}).$$

Returns

- **Proposition 1.** The net return is a linear approximation of the logarithmic return.
- **Proposition 2.** The *n*-period logarithmic return is given by

$$r_{n|0} = \log(P_n/P_0) = r_{1|0} + r_{2|1} + \cdots + r_{n|n-1}.$$

- **Proposition 3.** If prices are modeled as a Brownian motion in continuous time, the marginal distribution of prices is lognormal and the marginal distribution of log-returns is normal.

Financial Returns: stylized facts (Cont, 2001)

- Absence of autocorrelations
- Heavy tails
- Gain/loss asymmetry
- Aggregational Gaussianity
- Intermittency
- Volatility clustering
- Conditional heavy tails
- Slow decay of autocorrelation in absolute returns
- Leverage effect
- Volume/volatility correlation
- Asymmetry in time scales

Stochastic processes

- A stochastic process X_t is a (finite, countable or uncountable) family of random variables representing a statistical phenomenon that evolves in time according to probabilitic laws.
- The realization of a stochastic process is a path (a deterministic function of time).
- Given a stochastic process defined on some time horizon $(-T_1, T_2)$, if we fix some $t_0 \in (-T_1, T_2)$, we obtain a random variable X_{t_0}.
- The mean and variance functions of the process X_t are functions of time given by

$$\mu(t) = \mathrm{E}(X_t), \qquad t \in (-T_1, T_2)$$
$$\mathrm{var}(t) = \sigma^2(t) = \mathrm{var}(X_t), \qquad t \in (-T_1, T_2).$$

- Autocovariance function: covariance of X_{t_1} and X_{t_2}:

$$\gamma(X(t_1), X(t_2)) = \mathrm{E}\{[X_{t_1} - \mu(t_1)][X_{t_2} - \mu(t_2)]\}.$$

Stationary processes

- A stochastic process X_t is said to be **strictly stationary** if the joint distribution of $X_{t_1}, \ldots, X_{t_k}$ is the same as the joint distribution of $X_{t_1+\tau}, \ldots, X_{t_k+\tau}$.
- In other words, the joint distribution only depends on the intervals (lags) between $t_1, \ldots, t_k$, not on their location.
- Taking $k = 1$, it is easy to see that strict stationarity implies that the distribution of X_t is the same for all t, so that $\mathrm{E}(X_t) = \mu$ and $\mathrm{var}(X_t) = \sigma^2$ are constants.
- Taking $k = 2$, the autocovariance only depends on $\tau \stackrel{\text{def}}{=} t_2 - t_1$, so that the autocovariance function can be written as $\gamma(\tau) = \mathrm{E}[(X_t - \mu)(X_{t+\tau} - \mu)]$.
- Finally, the autocorrelation function is defined as $\rho(\tau) = \gamma(\tau)/\gamma(0) = \gamma(\tau)/\sigma^2$.

Weak stationarity

- A stochastic process X_t is said to be **weakly stationary** if $E(X_t) = \mu$ is constant and the autocovariance only depends on the lag $\tau \geq 0$: $\gamma(\tau) = E[(X_t - \mu)(X_{t+\tau} - \mu)]$.
- Properties of the autocorrelation function:
  1. $\rho(\tau) = \rho(-\tau)$;
  2. $|\rho(\tau)| \leq 1$;
  3. $\rho(\tau)$ does not uniquely identify the underlying process.

Basic models

- A **purely random** (or **white noise**) process Z_t is a sequence of iid (identically independently distributed) random variables with $\mathrm{E}(Z_t) = 0$ and $\mathrm{var}(Z_t) = \sigma^2$. Often it is also assumed $Z_t \sim N(0, \sigma_Z^2)$.
- By independence:

$$\gamma(k) = \mathrm{cov}(Z_t, Z_{t+k}) = \begin{cases} \sigma_Z^2 & k = 0; \\ 0 & k = \pm 1, \pm 2, \dots \end{cases}$$

- Suppose Z_t is a purely random process with $\mathrm{E}(Z_t) = \mu$ and $\mathrm{var}(Z_t) = \sigma^2$. A **random walk** process X_t is defined as

$$X_t = X_{t-1} + Z_t.$$

  Typically, $X_0 = 0$, so that

$$X_t = \sum_{i=1}^{t} Z_i.$$

Basic models

- Moreover

$$\mathrm{E}(X_t) = t\mu, \quad \mathrm{var}(X_t) = t\sigma_Z^2, \quad \gamma(\tau) = \tau\sigma_Z^2.$$

- Suppose Z_t is a white noise process with $\mathrm{var}(Z_t) = \sigma_Z^2$. X_t is said to be a **Moving Average** process of order q (MA(q)) if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}.$$

- It can be shown that

$$\mathrm{E}(X_t) = 0, \qquad \mathrm{var}(X_t) = \sigma_Z^2 \sum_{i=0}^{q} \beta_i^2$$

and

$$\gamma(k) = \begin{cases} 0 & k > q; \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & k = 1, \ldots, q; \\ \gamma(-k) & k < 0. \end{cases}$$

Basic models

- The autocorrelation function is given by

$$\rho(k) = \begin{cases} 1 & k = 0; \\ 0 & k > q; \\ \sum_{i=0}^{q-k} \beta_i \beta_{i+k} / \sum_{i=0}^{q} \beta_i^2 & k = 1, \ldots, q; \\ \rho(-k) & k < 0. \end{cases}$$

- For the MA(1) process with $\beta_0 = 1$:

$$\rho(k) = \begin{cases} 1 & k = 0; \\ \beta_1 / (1 + \beta_1^2) & k = \pm 1; \\ 0 & \text{otherwise.} \end{cases}$$

Invertibility

- A process X_t is said to be invertible if the random disturbance at time t (the so-called innovation) can be expressed as a convergent sum of previous and past values of X_t:

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

  where $\sum |\pi_j| < \infty$.
- An $MA(1)$ process $X_t = Z_t + \theta Z_{t-1}$ is invertible if $|\theta| < 1$.
- The *Backward shift* operator B is defined by

$$B^j X_t = X_{t-j}, \qquad \forall j,$$

  so that the $MA(q)$ process can be written as

$$\begin{aligned} X_t &= (\beta_0 + \beta_1 B + \cdots + \beta_q B^q) Z_t = \\ &= \theta(B) Z_t. \end{aligned}$$

Invertibility

- It can be shown that the process is invertible if the roots of the equation

$$\theta(B) = \beta_0 + \beta_1 B + \cdots + \beta_q B^q = 0$$

are larger than one in modulus.

- Example. For the $MA(1)$ process, we have

$$\theta(B) = 1 + \theta B = 0 \Leftrightarrow B = -\frac{1}{\theta},$$

whose modulus is larger than one if and only if $|\theta| < 1$.

- Thus, an $MA(1)$ process is invertible if and only if $|\theta| < 1$.

Autoregressive processes

- Let Z_t be a purely random process with $\mathrm{E}(Z_t) = 0$ and $\mathrm{var}(Z_t) = \sigma_Z^2$. X_t is an AutoRegressive process of order p ($AR(p)$) if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t. \tag{1}$$

- Consider the $AR(1)$ process $X_t = \alpha X_{t-1} + Z_t$. It is easy to verify that

$$X_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots ,$$

under the condition $-1 < \alpha < 1$. Using the backward shift operator B, equation (1) with $p = 1$ becomes

$$(1 - \alpha B) X_t = Z_t.$$

Autoregressive processes

- Using the fact that $\sum_{i=0}^{\infty} |\lambda|^i = 1/(1 - |\lambda|)$ when $|\lambda| < 1$, we get

$$\begin{aligned} X_t &= \frac{Z_t}{1 - \alpha B} = \\ &= (1 + \alpha B + \alpha^2 B^2 + \cdots) Z_t = \\ &= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots \end{aligned}$$

- From this expression we easily get $\mathrm{E}(X_t) = 0$, $\mathrm{var}(X_t) = \sigma_Z^2 (1 + \alpha^2 + \alpha^4 + \cdots)$. Thus, if $|\alpha| < 1$, $\mathrm{var}(X_t) = \sigma_X^2 = \sigma_Z^2 / (1 - \alpha^2)$.

Autoregressive processes

- The acv. f. is given by

$$
\begin{aligned}
\gamma(k) &= \mathrm{E}(X_t X_{t+k}) = \\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \alpha^i Z_{t-i}\right)\left(\sum_{j=0}^{\infty} \alpha^j Z_{t+k-j}\right)\right] = \\
&= \sigma_Z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} = \qquad \text{for } k \geq 0 \\
&= \frac{\alpha^k \sigma_Z^2}{1-\alpha^2} = \qquad \text{if} |\alpha| < 1 \\
&= \alpha^k \sigma_X^2.
\end{aligned}
$$

- Moreover, $\gamma(k) = \gamma(-k)$, so that, if $|\alpha| < 1$, the process is weakly stationary with ac. f. given by $\rho(k) = \alpha^{|k|}$ $(k = 0, 1, 2, \ldots)$.

Autoregressive processes

- In the general (p-th order) case, Equation (1) becomes

$$(1-\alpha_1 B-\cdots-\alpha_p B^p)X_t = Z_t \Leftrightarrow$$
$$X_t = \frac{Z_t}{(1-\alpha_1 B-\cdots-\alpha_p B^p)} =$$
$$= f(B)Z_t,$$

  where

$$f(B) = (1-\alpha_1 B-\cdots-\alpha_p B^p)^{-1} = (1+\beta_1 B+\beta_2 B^2+\cdots).$$

- However, finding the β_i's is not easy. Thus, to obtain the ac. f., the usual way consists in assuming stationarity, multiplying (1) by X_{t-k}, taking expectations and dividing by σ_X^2. This procedure gives the Yule-Walker equations

$$\rho(k) = \alpha_1\rho(k-1)+\alpha_2\rho(k-2)+\cdots+\alpha_p\rho(k-p), \qquad k>0.$$

Autoregressive processes

- Solution:

$$\rho(k) = A_1\pi_1^{|k|} + \cdots + A_p\pi_p^{|k|},$$

where the π_i's are the roots of $y^p - \alpha_1 y^{p-1} - \cdots - \alpha_p = 0$.

- The process is stationary if and only if $|\pi_i| < 1$ for all i or, equivalently, if the roots of

$$\phi(B) = 1 - \alpha_1 x - \cdots - \alpha_p x^p = 0$$

lie outside the unit circle.

ARMA processes

- An *ARMA*(p, q) model is a process containing p *AR* terms and q *MA* terms. It is given by

$$\begin{aligned} X_t = \mu + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \\ + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots + \beta_q Z_{t-q}. \end{aligned} \tag{2}$$

- Using the Backward shift operator we get:

$$\phi(B) X_t = \theta(B) Z_t,$$

where $\phi(B)$ and $\theta(B)$ are polynomials of order p and q:

$$\begin{aligned} \phi(B) &= 1 - \alpha_1 B - \cdots - \alpha_p B^p; \\ \theta(B) &= 1 + \beta_1 B + \cdots + \beta_q B^q. \end{aligned}$$

ARMA processes

- An $ARMA(p, q)$ process is stationary if the roots of $\phi(B) = 0$ lie outside the unit circle.
- An $ARMA(p, q)$ process is invertible if the roots of $\theta(B) = 0$ lie outside the unit circle.
- The coefficients of the pure MA representation $X_t = \psi(B)Z_t$ can be obtained as $\psi(B) = \theta(B)/\phi(B)$.
- The coefficients of the pure AR representation $\pi(B)X_t = Z_t$ can be obtained as $\pi(B) = \phi(B)/\theta(B)$.
- Thus, $\pi(B)\psi(B) = 1$.

ARIMA processes

- Problem: many observed time series are non-stationary.
- Thus, in order to fit a stationary $ARMA(p, q)$ model, one has to transform the data.
- Most common solution: differencing.
- If we write $W_t = \nabla^d X_t = (1 - B)^d X_t$, the $ARIMA(p, d, q)$ process is given by

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \cdots + \alpha_p W_{t-p} + \\ + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \cdots + \beta_q Z_{t-q},$$

or

$$\phi(B) W_t = \theta(B) Z_t \quad \text{or} \quad \phi(B)(1 - B)^d X_t = \theta(B) Z_t.$$

Fitting ARIMA models

- How do we fit a model to real data? There are two problems:
  1. choosing the model;
  2. estimating the parameters.
- As for the first issue, one might compare the sample autocovariance (autocorrelation) coefficient at lag k to the theoretical autocovariance (autocorrelation) function of a specific process.

Sample autocorrelation

- Given n observations $x_1, \dots, x_n$ from a time series X_t, there are $n-1$ pairs of observations separated by one time interval.
- The sample autocorrelation between X_t and X_{t+1} is

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{n-1}(x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{n-1}(x_{t+1} - \bar{x}_{(2)})^2}},$$

  where $\bar{x}_{(1)} = \sum_{t=1}^{n-1} x_t/(n-1)$ and $\bar{x}_{(2)} = \sum_{t=2}^{n} x_t/(n-1)$. $\hat{\rho}_1$ is called autocorrelation coefficient.
- It can be approximated as:

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\frac{n-1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2},$$

  where $\bar{x} = \sum_{t=1}^{n} x_t/n$.

Sample autocorrelation

- Most common approximation:

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.$$

- Autocorrelation coefficient at lag k:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.$$

- Autocovariance at lag k:

$$c_k = \frac{1}{n-k} \sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

It follows that

$$\hat{\rho}_k = \frac{c_k}{c_0}.$$

Sample autocorrelation and correlogram

- Notice that
$$c_0 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2$$
is just the variance of x_t.
- The correlogram is a scatterplot of k and r_k for some values of k (typically much smaller than n).
- Guidelines for interpretation:
  - for stationary series, $\hat{\rho}_1$ large and one or two further “large” (but smaller than $\hat{\rho}_1$) values;
  - alternating series: alternating correlogram;
  - for series with a trend: $\hat{\rho}_k > 0$ for many values of k;
  - seasonal data: correlogram exhibit oscillations at the same frequency.

Fitting ARIMA models

- It can be shown that, if $x_1, \ldots, x_n$ are iid observations from a distribution with arbitrary mean,

$$\hat{\rho}_k \overset{a}{\sim} N\left(-\frac{1}{n}, \frac{1}{n}\right).$$

- In general:
  - if the ac.f. cuts off at lag q, an $MA(q)$ process may be appropriate;
  - if the ac.f. decreases exponentially, an $AR(1)$ process may be appropriate;
  - other cases are more difficult to deal with.
- Estimating the mean is misleading if there are systematic components; even when there are no systematic components, the sample mean is often less informative than in classical statistics.

Fitting AR models

- Suppose the model has order p:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + Z_t.$$

- Given n observations we can estimate the parameters by least squares by minimizing

$$S = \sum_{t=p+1}^{n} [x_t - \mu - \alpha_1(x_{t-1} - \mu) - \cdots - \alpha_p(x_{t-p} - \mu)]^2.$$

- In the $AR(1)$ case the estimators are:

$$\hat{\mu} = \bar{x};$$

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n-1}(x_t - \bar{x})^2}. \tag{3}$$

Fitting AR models

- If we approximate the denominator of (3) with $\sum_{t=1}^{n}(x_t - \bar{x})^2$ we have $\hat{\alpha}_1 = r_1$.
- Asymptotically,

$$\mathrm{var}(\hat{\alpha}_1) = \frac{1 - \alpha_1^2}{n}.$$

- For the $AR(p)$ model we can either fit a regression model to

$$x_t - \bar{x} = \alpha_1(x_{t-1} - \bar{x}) + \alpha_2(x_{t-2} - \bar{x}) + \cdots + \alpha_p(x_{t-p} - \bar{x}) + Z_t$$

and use standard linear regression techniques, or insert the sample autocorrelations in the Yule-Walker equations and solve for $(\hat{\alpha}_1, \ldots, \hat{\alpha}_p)$.

Fitting AR models

- Possible tools for determining the order p of the process:
  - use the sample ac.f.;
  - use the partial ac.f. $\pi(k)$, defined as follows: when fitting an $AR(p)$ process, the last coefficient $\alpha_p = \pi(p)$ measures the *excess correlation* not accounted for by the first $p - 1$ parameters, namely by an $AR(p - 1)$ model. $\alpha_p = \pi(p)$ is the p-th partial autocorrelation coefficient. The partial autocorrelation function is the plot of p against α_p. It can be shown that the partial ac.f. of an $AR(p)$ process "cuts off" at lag p; moreover, it can be shown that $\pi(1) = \rho(1)$.
  - compute the residual sum of squares and plot it against p.

Fitting MA models

- In order to fit an *MA* model, as in the *AR* case, we have to:
  - find the order of the process;
  - estimate the parameters.
- As for the second problem, it is more difficult than the corresponding problem in the *AR* case, and numerical methods are needed.
- On the contrary, the first problem is easier, as it can usually be based on the sample ac.f., taking q equal to the largest value of k such that $\hat{\rho}_k$ is significantly different from zero.

Forecasting ARIMA models

- Having identified and estimated an appropriate ARIMA model, it is possible to use the ARIMA model equation directly. More precisely the forecast of X_{n+h} is obtained by replacing:
  - future values of Z by zero;
  - future values of X by their conditional expectation;
  - present and past values of X and Z by their observed values.
- A second strategy consists in using the ψ weights of the infinite MA representation, given by

$$X_{N+h} = Z_{N+h} + \psi_1 Z_{N+h-1} + \cdots$$

The forecast is then $\hat{X}_{n+h} = \sum_{j=0}^{\infty} \psi_{h+j} z_{N-j}$.

Forecasting ARIMA models

- The variance of the *h*-steps ahead forecast error is then

$$\begin{aligned}
&\mathrm{var}(X_{N+h} - \hat{X}_{N+h}) = \\
&= \mathrm{var}(Z_{N+h} + \psi_1 Z_{N+h-1} + \cdots + \psi_{h-1} Z_{N+1}) = \\
&= (1 + \psi_1^2 + \cdots + \psi_{h-1}^2)\sigma_Z^2. \qquad (4)
\end{aligned}$$

- A third strategy consists in using the π weights. We have

$$X_{N+h} = \pi_1 X_{N+h-1} + \pi_2 X_{N+h-2} + \cdots + \pi_h X_N + \cdots + Z_{N+h},$$

so that

$$\hat{X}_{N+h} = \pi_1 \hat{X}_{N+h-1} + \pi_2 \hat{X}_{N+h-2} + \cdots + \pi_h X_N + \pi_{h+1} X_{N-1} + \cdots$$

Prediction intervals

- Most P.I.s used in practice are of the form

$$\hat{X}_{N+h} \pm z_{\alpha/2}\sqrt{\text{var}(e_{N+h})},$$

  where $e_{N+h} = X_{N+h} - \hat{X}_{N+h}$ is the forecast error at time N and $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.
- It is implicitly assumed that $\hat{X}_{N+h}$ is an unbiased estimator of X_{N+h} and that e_{N+h} is normally distributed.
- In the Box-Jenkins approach, $\text{var}(e_{N+h})$ is given by (4), so that the P.I. is equal to

$$\hat{X}_{N+h} \pm z_{\alpha/2}\sqrt{(1 + \psi_1^2 + \cdots + \psi_{h-1}^2)\sigma_Z^2},$$

Models of financial returns

- Basic model:

$$r_t = \mu_t + \sigma_t \epsilon_t, \quad t = 1, 2, \dots, \quad \epsilon_t \sim iid\ WN(0, 1), \qquad (5)$$

  where *WN* stands for *white noise*.
- If μ_t is an ARMA model and $\sigma_t = \sigma\ \forall t$, then r_t is an ARMA model. Not realistic. . .
- . . . because the variance is usually NOT constant.
- Hence: need to model a time-varying variance σ_t^2.
- Moreover: the mean is typically small and almost constant, so that setting $\mu_t = \mu\ \forall t$ makes little difference.
- Conclusion: modeling the variance is the real challenge.

GARCH models

- Let's go back to (5), assuming that $\mu_t = \mu$ and define r'_t as the time-t residual return, i.e. $r'_t = r_t - \mu$. In the following we will model r'_t but, with a slight abuse of notation, we will call it r_t.
- ARIMA models assume that the variance is constant. If it is not, we need to model a time-varying variance.
- A GARCH(m, s) model for the residual returns is:

$$r_t = \sigma_t \epsilon_t, \quad t = 1, 2, \dots, \quad \epsilon_t \sim iid\ WN(0, 1),$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i) < 1$.

The ARCH LM test

- A formal way of checking whether an ARCH model should be used is Engle's Lagrange multiplier (LM) test, which is based on the the following steps:
  1. Estimate the best-fitting AR(q) model and compute the residuals $\hat{\epsilon}_t$;
  2. fit the regression $\hat{\epsilon}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^q \hat{\alpha}_i \hat{\epsilon}_{t-i}^2$;
  3. under $H_0 : \alpha_i = 0$ ($i = 1, \ldots, q$), the random variable $(T - q)R^2$ is distributed as χ_q^2, where T is the number of observations.
- Another possibility is the usual Ljung-Box test applied to the standardized squared residuals.

The GARCH$(1, 1)$ model

- In this case, $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. Hence:
  - A large r_{t-1}^2 or σ_{t-1}^2 tends to give a large σ_t^2;
  - The unconditional variance is

$$\text{var}(r_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1};$$

  - If $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, the kurtosis is given by

$$\kappa = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

- Any GARCH process has unconditional mean equal to 0 and is serially uncorrelated. Hence, if the variance exists, it is weakly stationary;
- Any GARCH process can be written as an infinite order ARCH process.

GARCH prediction

- It can be shown that the h-step ahead forecast of a GARCH(1,1) process is given by

$$\sigma_{t+h}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{t+h-1}^2, \quad t > 1, \tag{6}$$

  which can be rewritten as

$$\sigma_{t+h}^2 = \frac{\alpha_0[1 - (\alpha_1 + \beta_1)^{h-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{h-1}\sigma_{t+1}^2,$$

  so that

$$\sigma_{t+h}^2 \overset{h\to\infty}{\longrightarrow} \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

Extensions of the basic GARCH model: 1. IGARCH

- Let

$$r_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) r_{t-1}^2,$$

  with $\beta_1 \in (0, 1)$. This is called an Integrated GARCH (IGARCH) model.
- It is nonstationary with infinite variance.
- The special case obtained when $\alpha_0 = 0$ is the Exponentially weighted moving average (EWMA) model, popularized by JP Morgan's RiskMetrics™ approach to market risk measurement.
- Using (6) with $\alpha_1 + \beta_1 = 1$, the *h*-step ahead forecast is

$$\sigma_{t+h}^2 = \sigma_{t+1}^2 + (h - 1)\alpha_0, \quad t > 1,$$

Extensions of the basic GARCH model: 2. GARCH-M

- Let

$$r_t = \mu + c\sigma_t^2 + \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta_1 r_{t-1}^2,$$

  with $\alpha_1, \beta_1 \in (0, 1)$, $\alpha_1 + \beta_1 < 1$. This is called a GARCH-in-mean (GARCH-M) model.
- c can be interpreted as a risk premium parameter.
- The process r_t is serially correlated.

Extensions of the basic GARCH model: 3. EGARCH

- Let
$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - \mathrm{E}(|\epsilon_t|)], \quad \theta, \gamma \in \mathbb{R}.$$
- An EGARCH(m, s) model is given by
$$r_t = \sigma_t\epsilon_t, \quad \log(\sigma_t^2) = \frac{1 + \beta_1 B + \cdots + \beta_s B^s}{1 - \alpha_1 B - \cdots - \alpha_m B^m} g(\epsilon_{t-1}),$$
- $g(\epsilon_t)$ is asymmetric, so that the model gives, in general, different weight to positive and negative shocks.

Evaluation of models

- As the variance of an asset return is not directly observable, comparing the forecasting performance of different volatility models is difficult.
- It is preferable to use residual analysis or information criteria.
- Akaike's Information Criterion (AIC):

$$AIC = \frac{-2\log(L)}{n} + \frac{2m}{n},$$

- Bayesian Information Criterion (BIC):

$$BIC = \frac{-2\log(L)}{n} + \frac{m\log(n)}{n},$$

where m is the number of parameters and $\log(L)$ is the maximized log-likelihood.

Value-at-Risk

- Suppose that at time t we are interested in the risk of a financial position for the next h periods. Let $\Delta V(h)$ be the change in value of the assets in the financial position from t to $t+h$.
- Let $F_h(x)$ be the cdf of $\Delta V(h)$. The p-level $(0 < p < 1)$ Value-at-Risk (VaR) over the time horizon h is defined as follows:

$$VaR_p : P(\Delta V(h) \leq VaR_p) = F_h(VaR_p) = p.$$

RiskMetrics VaR

- Let $\mathcal{F}_t$ be the information available at time t. Let $r_t|\mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$.
- Let $\mu_t = 0$, $\sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)r_{t-1}^2$.
- This is an IGARCH(1,1) model without drift.
- Let $r_t[k] = r_{t+1} + \cdots + r_{t+k-1} + r_{t+k}$. It can be shown that $r_t[k]|\mathcal{F}_t \sim N(0, k\sigma_{t+1}^2)$.
- Conditional k-period VaR at level p:

$$VaR_p(k) = \text{amount of position} \times \sqrt{k} \times z_p \times \sigma_{t+1}.$$

- RiskMetrics sets $\alpha = 0.94$.

One-period GARCH(m, s) VaR

- Let the model be:

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j a_{t-j},$$

$$a_t = \sigma_t \epsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2.$$

- One-step ahead forecasts:

$$\hat{r}_{t+1} = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t+1-i} + \sum_{j=1}^{q} \theta_j a_{t+1-j},$$

$$\hat{\sigma}_{t+1}^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t+1-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t+1-j}^2.$$

One-period VaR

- If $\epsilon_t \sim N(0,1)$, $r_{t+1}|\mathcal{F}_t \sim N(\hat{r}_{t+1}, \hat{\sigma}^2_{t+1})$ and the p-level VaR is
$$VaR_p = \hat{r}_{t+1} + z_p\hat{\sigma}_{t+1}.$$
- If $\epsilon_t \sim t^*(\nu)$ (the **standardized** t distribution with ν degrees of freedom), $r_{t+1}|\mathcal{F}_t \sim N(\hat{r}_{t+1}, \hat{\sigma}^2_{t+1})$ and the p-level VaR is
$$VaR_p = \hat{r}_{t+1} + t^*_{p,\nu}\hat{\sigma}_{t+1},$$
where $t^*_{p,\nu}$ is the p-quantile of the standardized t distribution with ν degrees of freedom.
- Note that $t^*_{p,\nu} = t_{p,\nu}/\sqrt{\nu/(\nu-2)}$, provided that $\nu > 2$.

Backtesting

- Statistical procedure where actual profits and losses are systematically compared to corresponding VaR estimates.
- For example, if the VaR level is 95%, we expect, on average, an exception in every 20 days.
- Tests of unconditional coverage: check whether the frequency of exceptions over some specified time interval is in line with the VaR level.
- Tests of conditional coverage: check whether the frequency of exceptions over some specified time interval is in line with the VaR level **and** exceptions are evenly spread over time.
- See Jorion (2006) for details.

References

Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1:223–236

Jorion P (2006) Value at Risk: The New Benchmark for Measuring Financial Risk, 3rd edn. Wiley