# Time Series Analysis

## 2. Non-stationary univariate time series

Andrew Lesniewski

Baruch College
New York

Fall 2019

# Outline

# Trend-stationary and unit root processes

- So far we have focused on time series that are stationary (or, more precisely, covariance-stationary).
- We have seen that a stationary time series in the $ARMA(p, q)$ family can be written in the moving average (MA) form:

$$\begin{aligned} X_t &= \mu + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} \dots \\ &= \mu + \gamma(L) \varepsilon_t , \end{aligned} \tag{1}$$

  where $L$ is the lag operator, and where $\sum_{j=1}^{\infty} |\gamma_j| < \infty$.
- Stationary series are rather unusual in finance, and hence the need for developing models that capture the non-stationary nature of financial time series.
- There are various approaches to model non-stationarity. We will initially focus on two of them:
  (i) Non-stationary process with a deterministic trend and stationary disturbances.
  (ii) Non-stationary process with a unit root (non-stationary disturbances).

# Trend-stationary and unit root processes

- An example of the former type of a time series is the following process with *linear trend*:

$$X_t = \alpha + \delta t + \gamma(L)\varepsilon_t, \tag{2}$$

where $\alpha, \delta \in \mathbb{R}$. This amounts to replacing the constant mean $\mu$ of the stationary process (1) with a linear function.

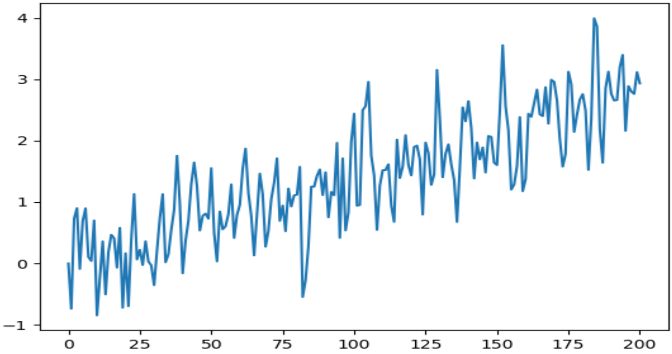
- The process behaves thus like a pure deterministic trend perturbed by a stationary random noise, and is referred to as a *trend-stationary* process.
- This is to be contrasted with the second type of non-stationarity mentioned above.
- Consider the following *unit root* process:

$$X_t = \alpha + X_{t-1} + \gamma(L)\varepsilon_t. \tag{3}$$

Here the non-stationarity comes from the presence of the unit root $\beta = 1$ in the autoregressive part of the specification above.

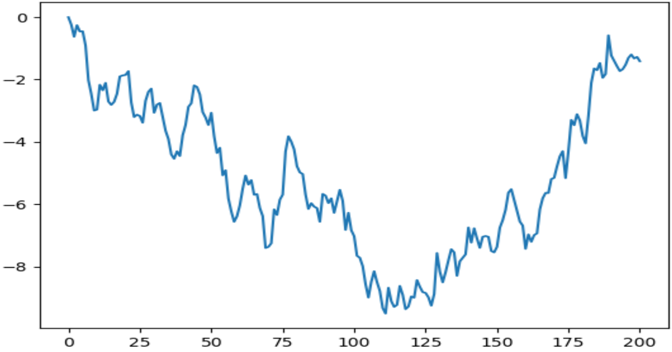
# Trend-stationary and unit root processes

- The graph below shows a simulated $AR(1)$ time series (2) (with deterministic trend $\alpha + \delta t$) with the following choice of parameters: $\alpha = 0.0$, $\delta = 0.01$, $\beta = 0.3$, $\sigma = 0.5$.



# Trend-stationary and unit root processes

- The graph below shows a simulated unit root $AR(1)$ time series with the following choice of parameters: $\alpha = 0.0$, $\beta = 1.0$, $\sigma = 0.5$.



# Trend-stationary and unit root processes

- A unit root process exhibits purely stochastic, persisting trends that have their source in the non-stationarity of the shocks delivered to the system.
- Iterating (3), we can write it as

$$X_t = X_0 + \alpha t + \gamma(L) \sum_{j=0}^{t} L^j \varepsilon_t . \tag{4}$$

- This representation shows explicitly that the variance of the random shock grows linearly in $t$.
- Processes of this form are also referred to as *integrated* of order 1 and are denoted by $I(1)$.
- This name is motivated by the following observation.

# Trend-stationary and unit root processes

- The operator
$$\Delta = 1 - L \tag{5}$$
is called the first *difference operator*.
- One can think about $\Delta$ as a discretized time derivative. Indeed,
$$\begin{aligned}\Delta X_t &= (1 - L)X_t \\ &= X_t - X_{t-1}.\end{aligned}$$
- Using it, we can write (3) in the form:
$$\Delta X_t = \alpha + \gamma(L)\varepsilon_t. \tag{6}$$
- Equation (4) is then the "integrated" version of this difference equation.

# ARIMA models

- More generally, integrated processes $I(d)$ of any *integer* order $d \geq 1$ are of the form:
$$\Delta^d X_t = \omega + \gamma(L)\varepsilon_t. \tag{7}$$
- Here, $\Delta^d$ denotes the $d$-th power of the difference operator $\Delta$.
- For example, for $d = 2$, $(1 - L)^2 = 1 - 2L + L^2$, and an $I(2)$ process can be written as
$$X_t = \omega + 2X_{t-1} - X_{t-2} + \gamma(L)\varepsilon_t. \tag{8}$$
- We will now recast these ideas in terms of ARIMA models.

# ARIMA models

- Recall that an $ARMA(p, q)$ model can be written in the form:

$$\psi(L)X_t = \alpha + \varphi(L)\varepsilon_t, \tag{9}$$

where

$$\begin{aligned} \psi(z) &= 1 - \beta_1 z - \ldots - \beta_p z^p, \\ \varphi(z) &= 1 + \theta_1 z + \ldots + \theta_q z^q. \end{aligned} \tag{10}$$

Covariance-stationarity requires that the roots of $\psi(z)$ lie outside of the unit circle.

- This coincides with equation (1), if we set $\mu = \alpha/\psi(1)$ and $\gamma(L) = \psi(L)^{-1}\varphi(L)$.
- We will now assume that the characteristic polynomial has a unit root of degree $d > 0$, i.e. it is of the form $\psi(z)(1 - z)^d$, where $\psi(z)$ is a polynomial of degree $p$ with roots outside of the unit circle.

# ARIMA models

- This leads us to the concept of an *autoregressive integrated moving average (ARIMA) model.*
- An *ARIMA*$(p, d, q)$ model is specified as follows:

$$\psi(L)(1-L)^d X_t = \alpha + \varphi(L)\varepsilon_t. \tag{11}$$

- Here, $p$ is the number of autoregressive lags (without the unit roots), $d$ is the order of integration (the order of the unit root), and $q$ is the number of the moving average lags.
- Equivalently, the specification of an *ARIMA*$(p, d, q)$ time series can be written as

$$(1-L)^d X_t = \mu + \psi(L)^{-1}\varphi(L)\varepsilon_t, \tag{12}$$

  where $\mu = \alpha/\psi(1)$.
- Python implementation of *ARIMA*$(p, d, q)$ is in the package `statsmodels`.

# ARIMA models

- Examples of $ARIMA(p, d, q)$ models include:
  (i) $ARIMA(0, 0, 0)$. This is simply the white noise process:
  $$X_t = \varepsilon_t. \tag{13}$$
  (ii) $ARIMA(0, 1, 0)$. This is the random walk process:
  $$X_t = X_{t-1} + \varepsilon_t. \tag{14}$$
  (iii) $ARIMA(1, 0, 1)$. This is the *exponentially weighted moving average model* (EWMA):
  $$X_t = \lambda X_{t-1} + \varepsilon_t + (1 - \lambda)\varepsilon_{t-1}. \tag{15}$$
  (iv) $ARIMA(0, 2, 2)$. This is a general linear exponential smoothing model:
  $$X_t = 2X_{t-1} - X_{t-2} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}. \tag{16}$$
  In this model both the level and the slope of the time series are smoothed using exponentially weighted moving averages.

# Forecasting non-stationary time series

- Forecasting a non-stationary time series uses the methodology explained in Lecture Notes #1. Consider first the trend stationary time series (2).
- A one-period forecast is given by

$$\begin{aligned} X^*_{t+1|1:t} &= \mathrm{E}_t(X_{t+1}) \\ &= \alpha + \delta(t+1) + \gamma_1\varepsilon_t + \gamma_2\varepsilon_{t-1} + \ldots, \end{aligned} \tag{17}$$

  since $\mathrm{E}_t(\varepsilon_{t+1}) = 0$.
- Likewise, a $k$-period forecast is given by

$$\begin{aligned} X^*_{t+k|1:t} &= \mathrm{E}_t(X_{t+k}) \\ &= \alpha + \delta(t+k) + \gamma_k\varepsilon_t + \gamma_{k+1}\varepsilon_{t-1} + \ldots, \end{aligned} \tag{18}$$

  since $\mathrm{E}_t(\varepsilon_{t+j}) = 0$, for all $j > 0$.

# Forecasting non-stationary time series

- Let us now estimate the magnitude of the forecast error. Its value is

$$\begin{aligned} X_{t+k} - X^*_{t+k|1:t} &= \alpha(t+k)\delta + \varepsilon_{t+k} + \gamma_1\varepsilon_{t+k-1} + \gamma_2\varepsilon_{t+k-2} + \dots \\ &\quad + \gamma_{k-1}\varepsilon_{t+1} + \gamma_k\varepsilon_t + \gamma_{k+1}\varepsilon_{t-1} + \dots \\ &\quad - \big(\alpha + \delta(t+k) + \gamma_k\varepsilon_t + \gamma_{k+1}\varepsilon_{t-1} + \dots\big) \\ &= \varepsilon_{t+k} + \gamma_1\varepsilon_{t+k-1} + \gamma_2\varepsilon_{t+k-2} + \dots + \gamma_{k-1}\varepsilon_{t+1}. \end{aligned}$$

- The variance of the forecast error is

$$\mathrm{E}_t\big((X_{t+k} - X^*_{t+k|1:t})^2\big) = \sigma^2(1 + \gamma_1^2 + \dots + \gamma_{k-1}^2). \tag{19}$$

- Note that the series on the RHS converges as the time horizon $k$ goes to infinity, and its limit is the variance of $\gamma(L)\varepsilon_t$.

# Forecasting non-stationary time series

- For the unit root process (3), the one-period forecast is

$$X^*_{t+1|1:t} = X_t + \delta + \gamma_1 \varepsilon_t + \gamma_2 \varepsilon_{t-1} + \ldots . \tag{20}$$

- More generally, the $k$-period forecast is

$$X^*_{t+k|1:t} = X_t + \delta k + (\gamma_1 + \ldots + \gamma_k)\varepsilon_t + (\gamma_2 + \ldots + \gamma_{k+1})\varepsilon_{t-1} + \ldots . \tag{21}$$

- It is easy to see that the forecast error is

$$\begin{aligned} X_{t+k} - X^*_{t+k|1:t} = \varepsilon_{t+k} + (1+\gamma_1)\varepsilon_{t+k-1} + (1+\gamma_1+\gamma_2)\varepsilon_{t+k-2} + \ldots \\ + (1+\gamma_1+\ldots+\gamma_{k-1})\varepsilon_{t+1} . \end{aligned} \tag{22}$$

# Forecasting non-stationary time series

- The variance of the forecast error is

$$\mathrm{E}_t\big((X_{t+k} - X^*_{t+k|1:t})^2\big) = \sigma^2(1 + (1+\gamma_1)^2 + \ldots + (1 + \gamma_1 + \ldots + \gamma_{k-1})^2). \quad (23)$$

- The quality of the forecast deteriorates significantly with the length of the forecasting horizon: this expression diverges linearly (proportionally to $k$), as $k \longrightarrow \infty$.
- In summary, for a trend-stationary process the forecast error remains bounded as the forecasting horizon increases.
- In contrast, for a unit root process, the forecast error increases (asymptotically) linearly, as the length of the horizon goes to infinity.

# Dickey-Fuller test

- It is of practical importance to determine whether a time series has a unit root. A number of statistical tests for detecting the presence of a unit root in a time series.
- The *Dickey-Fuller test* (DF) tests the null hypothesis $H_0$ of whether a unit root is present in an $AR(1)$ model against the alternative hypothesis $H_a$ of stationarity.
- In other words,

$$H_0 : \beta = 1, \quad \text{against } H_a : \beta < 1. \tag{24}$$

- Various versions of this test address different model specifications.
- Theoretically, this test requires the knowledge of the probability distribution of the estimated coefficient $\widehat{\beta}$. We will heuristically explain how to determine this probability distribution in the large $T$ limit.

# Dickey-Fuller test

- Consider first the case of the $AR(1)$ time series with $\alpha = 0$, $\delta = 0$. As we saw in Lecture Notes #1, the MLE estimate of $\beta$ is given by

$$\widehat{\beta} = \frac{\sum_{t=0}^{T-1} x_t x_{t+1}}{\sum_{t=0}^{T-1} x_t^2}\,. \tag{25}$$

- The deviation of $\widehat{\beta}$ from 1 can be thus expressed as follows:

$$T(\widehat{\beta} - 1) = \frac{\frac{1}{T}\sum_{t=0}^{T-1} x_t \widehat{\varepsilon}_{t+1}}{\frac{1}{T^2}\sum_{t=0}^{T-1} x_t^2}\,. \tag{26}$$

- Let $W(s)$, $0 \le s \le 1$, denote the standard Brownian motion.

# Dickey-Fuller test

- Note that, under the null hypothesis $H_0$, $\frac{1}{\sqrt{T}}\widehat{\varepsilon}_{t+1}$ can be thought of as the increment of $W(s)$ from $\frac{t}{T}$ to $\frac{t+1}{T}$, multiplied by $\sigma$:

$$\frac{1}{\sqrt{T}}\,\widehat{\varepsilon}_{t+1} = \sigma\Big(W\big(\tfrac{t+1}{T}\big) - W\big(\tfrac{t}{T}\big)\Big).$$

- Therefore, if the true value of $\beta$ is 1, then as $T \longrightarrow \infty$,

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1} x_t\widehat{\varepsilon}_{t+1} &\longrightarrow \sigma^2\int_0^1 W(s)dW(s)\\ &= \frac{1}{2}\,\sigma^2(W(1)^2-1).\end{aligned}$$

- Similarly,

$$\frac{1}{T^2}\sum_{t=0}^{T-1} x_t^2 \longrightarrow \sigma^2\int_0^1 W(s)^2 ds.$$

# Dickey-Fuller test

- As a result, under the null hypothesis $H_0$,

$$T(\widehat{\beta}-1) \longrightarrow \frac{1}{2}\,\frac{W(1)^2-1}{\int_0^1 W(s)^2 ds}\,. \tag{27}$$

- The variable $T(\widehat{\beta}-1)$ describes asymptotically the deviation of $\widehat{\beta}$ from 1. It is the ratio of a $\chi^2$-distributed variable and a non-standard random variable.
- Unfortunately, the distribution of the random variable on the RHS of the expression above is not known in closed form!
- In practice, critical values of this random variables can be calculated for finite values of $T$ by means of Monte Carlo simulations for a number of benchmark levels. Their values are stored in computer software.

# Dickey-Fuller test

- Alternatively, one can use the Dickey-Fuller $t$-stat:

$$t_\beta = \frac{\widehat{\beta} - 1}{\widehat{\sigma}_\beta}\,, \tag{28}$$

where

$$\widehat{\sigma}_\beta^2 = \widehat{\sigma}^2 + \sum_{t=1}^{T} x_{t-1}^2 \tag{29}$$

is the estimated variance for $\widehat{\beta} - 1$, and $\widehat{\sigma}^2$ is the estimated variance of the residuals.

- The limiting distribution of $t_\beta$ is not Gaussian. Reasoning as above, we see that its limit is explicitly given by

$$t_\beta \longrightarrow \frac{1}{2}\,\frac{W(1)^2 - 1}{\sqrt{\int_0^1 W(s)^2 ds}}\,. \tag{30}$$

# Dickey-Fuller test

- Next, we consider the same $AR(1)$ dynamics without drift, but we this time we include estimated $\alpha$ in the test. In this case, the large $T$ limits are:

$$\begin{aligned} T(\widehat{\beta}-1) &\longrightarrow \frac{\frac{1}{2}\left(W(1)^2-1\right)-W(1)\int_0^1 W(s)ds}{\int_0^1 W(s)^2 ds-\left(\int_0^1 W(s)ds\right)^2}\,, \\ \sqrt{T}\widehat{\alpha} &\longrightarrow \sigma\,\frac{W(1)\int_0^1 W(s)^2 ds-\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(s)^2 ds-\left(\int_0^1 W(s)ds\right)^2}\,. \end{aligned} \tag{31}$$

- Again, neither $T(\widehat{\beta}-1)$ nor $\sqrt{T}\widehat{\alpha}$ are normally distributed in the limit of large $T$.
- Their critical values, for finite $T$, can again be found by means of Monte Carlo simulations.

# Dickey-Fuller test

- Alternatively, one can use the Dickey-Fuller $t$-stat defined in analogy to (28). Its $T \to \infty$ limit is

$$t_\beta \longrightarrow \frac{\frac{1}{2}\left(W(1)^2 - 1\right) - W(1)\int_0^1 W(s)ds}{\sqrt{\int_0^1 W(s)^2 ds - \left(\int_0^1 W(s)ds\right)^2}} . \tag{32}$$

- The limit distribution is non-Gaussian: the random variable on the right hand side of the expression above is the ratio of two non-standard variables.
- The technical details of the informal limit arguments can be found in Chapter 17 of Hamilton's book [1].
- Finite $T$ critical values are calculated by means of Monte Carlo simulations, and are available in software packages.
- There are a number of other cases (true process having a constant term, or a deterministic drift), for which DF tests have been developed (see, again [1]).

# Augmented Dickey-Fuller test

- A useful extension of the DF test is the *augmented Dickey-Fuller test* (ADF). It is designed to test the null hypothesis of the presence of a unit root in a general $AR(p)$ time series sample against the alternative hypothesis of stationarity.
- Let us discuss the simplest case of an $AR(p)$ process whose true dynamics does not have the constant term:

$$(1 - \beta_1 L - \ldots - \beta_p L^p) X_t = \varepsilon_t. \tag{33}$$

- Define

$$\begin{aligned} \beta &= \beta_1 + \ldots + \beta_p, \\ \gamma_j &= -(\beta_{j+1} + \ldots + \beta_p), \quad \text{for } j = 1, \ldots, p-1. \end{aligned} \tag{34}$$

- After some algebra we can then write (33) in the form:

$$\big(1 - \beta L - (\gamma_1 L + \ldots + \gamma_{p-1} L^{p-1})(1 - L)\big) X_t = \varepsilon_t. \tag{35}$$

# Augmented Dickey-Fuller test

- For example, in the case of $p = 2$, the algebra goes as follows:

$$\begin{aligned}1 - \beta_1 L - \beta_2 L^2 &= 1 - (\beta_1 + \beta_2)L + \beta_2 L - \beta_2 L^2 \\ &= 1 - (\beta_1 + \beta_2)L + \beta_2 L(1 - L) \\ &= 1 - \beta L - \gamma_1 L(1 - L),\end{aligned}$$

  as $\beta_1 + \beta_2 = \beta$ and $-\beta_2 = \gamma_1$.
- As a consequence,

$$\big(1 - \beta L - \gamma_1 L(1 - L)\big) X_t = \varepsilon_t .$$

  The general case (35) follows in an analogous manner.
- Equivalently, equation (35) can be written as

$$X_t = \beta X_{t-1} + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_{p-1} \Delta X_{t-p+1} + \varepsilon_t . \qquad (36)$$

  This form of the $AR(p)$ process is sometimes referred to as the *error correcting model* (ECM).

# Augmented Dickey-Fuller test

- Suppose now that $X_t$ has a single unit root. This means that $\psi(1) = 0$, and so $\beta = 1$, while other roots of $\psi(z)$ lie outside of the unit circle.
- We can then write

$$\psi(z) = (1 - z)(1 - \gamma_1 z - \ldots - \gamma_{p-1} z^{p-1}). \tag{37}$$

- As a consequence, under the null hypothesis $\beta = 1$, the dynamics (35) of $X_t$ can be written as

$$(1 - \gamma_1 L - \ldots - \gamma_{p-1} L^{p-1})(1 - L) X_t = \varepsilon_t .$$

  Multiplying both sides of this equation by $(1 - \gamma_1 L - \ldots - \gamma_{p-1} L^{p-1})^{-1}$, we find that

$$X_t = X_{t-1} + (1 - \gamma_1 L - \ldots - \gamma_{p-1} L^{p-1})^{-1} \varepsilon_t . \tag{38}$$

# Augmented Dickey-Fuller test

- This has almost the same form as the $AR(1)$ model without a constant term!
- The difference is that the shocks $\varepsilon_t$ are replaced with more complex (but still stationary) shocks $z_t = (1 - \gamma_1 L - \ldots - \gamma_{p-1} L^{p-1})^{-1} \varepsilon_t$.
- The calculations carried out above for $p = 1$ can be generalized to the general case by taking into account the statistics of the shocks $z_t$.
- The augmented Dickey-Fuller test is a test on the $t$-stat $t_\beta$ defined in analogy to the univariate case.
- A thorough discussion of the ADF test is presented in Chapter 17 of [1].
- The test is implemented as the function `adfuller` in `statsmodels`.

# ARIMA models

- Unit root time series tend to be finicky, and the presence of a unit root may be a transient phenomenon.
- A common cause of breakdown of stationarity is a *structural break*: an unexpected shift in the time series, often caused by an exogenous shock.
- The presence of a structural break can create the appearance that the shocks have permanent, rather than transitory, effects. That may bias the conclusion of a test towards a unit root.
- Structural breaks can lead to model misspecification and, consequently, significant forecasting errors.

# ARFIMA models

- An *autoregressive fractionally integrated moving average model* $ARFIMA(p, d, q)$ is an extension of the $ARIMA(p, d, q)$ model to non-integer $d$.
- The fractional power of $1 - L$ is *defined* as

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \\ &= \sum_{k=0}^{\infty} \frac{d(d-1)\dots(d-k+1)}{k!} (-L)^k \\ &= 1 - dL + \frac{1}{2} d(d-1)L^2 + \dots .\end{aligned} \tag{39}$$

- Note that, for fractional $d$, this is an infinite power series in $L$ (rather than a polynomial of order $d$, if $d$ is integer).
- An $ARFIMA(p, d, q)$ model is specified by the same equation as an $ARIMA(p, d, q)$ model:

$$\psi(L)(1-L)^d X_t = \alpha + \varphi(L)\varepsilon_t . \tag{40}$$

# ARFIMA models

- As an example, consider a simple model $ARFIMA(0, d, 0)$ specified by:

$$(1 - L)^d X_t = \varepsilon_t, \tag{41}$$

or

$$X_t = dX_{t-1} - \frac{1}{2} d(d-1)X_{t-2} + \ldots + \varepsilon_t. \tag{42}$$

- Alternatively, using

$$(1 - L)^{-d} = 1 + dL + \frac{1}{2} d(d+1)L^2 + \ldots,$$

we can write this in the moving average form:

$$X_t = \varepsilon_t + d\varepsilon_{t-1} + \frac{1}{2} d(d+1)\varepsilon_{t-2} + \ldots. \tag{43}$$

# ARFIMA models

- There is no simple closed form expression for the ACF in this model.
- However, with a bit of work one can show that, asymptotically as $j \to \infty$, the ACF has a power law behavior,

$$\begin{aligned} R_j &\approx |j|^{2d-1} \\ &= |j|^{2H}, \end{aligned} \tag{44}$$

  where the *Hurst exponent* $H$ is given by $H = d - 1/2$.
- This is similar to the properties of the *fractional Brownian motion* $W_H(t)$.
- In particular, if $H < 0$, the ACF decays as a power of $|j|$.
- This should be contrasted with the exponential (or even faster) decay of autocorrelations in the *ARMA* models.
- *ARFIMA*$(p, d, q)$ models are thus useful for modeling time series with *long memory*.

# Description of cointegration

- Various economic / financial series exhibit the phenomenon of *cointegration*.
- Two time series $X_t$ and $Y_t$ are said to be *cointegrated* if
  (i) Each of the series is $I(1)$, i.e. each time series is unit root non-stationary.
  (ii) Some linear combination of $X_t$ and $Y_t$ is covariance-stationary.
- A vector $a$, such that the time series

$$a^{\mathrm{T}} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = a_1 X_t + a_2 Y_t \tag{45}$$

  is stationary, is called a *cointegrating vector*.
- It may also be useful to consider more than two time series. In this case, one can talk of one or more cointegrating vectors.
- From the economic perspective, cointegration between two time series means that there is a long-term equilibrium relationship between the two time series, even though they are exposed to random shocks during their time evolution.

# Cointegrated series

- As an example, we consider the following two time series:

$$\begin{aligned} X_t &= \alpha_1 + \gamma Y_t + \varepsilon_{t,1}, \\ Y_t &= \alpha_2 + Y_{t-1} + \varepsilon_{t,2}, \end{aligned} \tag{46}$$

  where the innovations $\varepsilon_{t,1} \sim N(0, \sigma_1^2)$ and $\varepsilon_{t,2} \sim N(0, \sigma_2^2)$ are independent.
- Both of these series are $I(1)$. Indeed,

$$\begin{aligned} \Delta Y_t &= \alpha_2 + \varepsilon_{t,2}, \\ \Delta X_t &= \gamma \Delta Y_t + \varepsilon_{t,1} - \varepsilon_{t-1,1} \\ &= \gamma\alpha_2 + \gamma\varepsilon_{t,2} + \varepsilon_{t,1} - \varepsilon_{t-1,1}, \end{aligned}$$

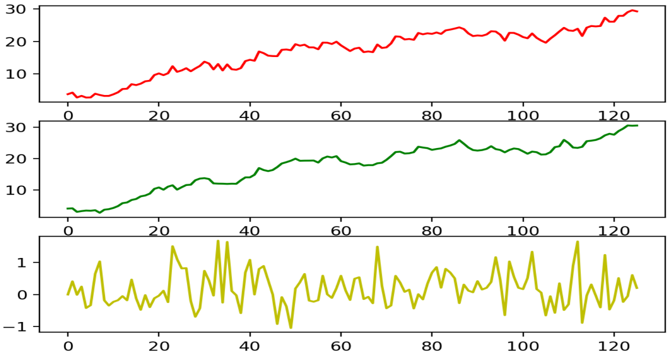
  and so both differenced processes $\Delta X_t$ and $\Delta Y_t$ are stationary.
- The process

$$X_t - \gamma Y_t = a^{\mathrm{T}} \begin{pmatrix} X_t \\ Y_t \end{pmatrix}, \quad \text{where } a = \begin{pmatrix} 1 \\ -\gamma \end{pmatrix}, \tag{47}$$

  is stationary, and thus the vector $a$ is a cointegrating vector for $X_t$ and $Y_t$.

# Cointegrated series

- The graph below shows a simulation of the system (46) with the following choice of parameters: $\alpha_1 = 0.1$, $\alpha_2 = 0.2$ $\gamma = 0.95$, $\sigma_1 = 0.5$, $\sigma_2 = 0.7$. The top graph shows $X_t$, the middle graph shows $Y_t$, and the bottom graph shows $X_t - \gamma Y_t$.



# Testing for cointegration

- We will now discuss statistical tests for cointegration. We continue assuming two univariate time series; we will address the general case in Lecture Notes #3.
- The test discussed below was originally developed by Engle and Granger, and it is based on regression techniques.
- Let $x_t$ and $y_t$ be the observations of the two time series, whose cointegration we want to test. We proceed as follows:
  - (i) We first run the DF (or ADF) test for each of the series $x_t$ and $y_t$ is $I(1)$.
  - (ii) Assuming, they both pass, we run the OLS regression $x = \gamma y + \alpha$, and form the vector $a$ as in formula (47).
  - (iii) We now test whether $a$ is a cointegrating vector for $x_t$ and $y_t$. To this end, we first calculate the "cointegrating residual process"

$$u_t = a^{\mathrm{T}} \begin{pmatrix} x_t \\ y_t \end{pmatrix} . \tag{48}$$

  - (iv) We perform a unit root test on $u_t$ to determine whether it is $I(0)$ (i.e. stationary).

# Testing for cointegration

- Notice that step (ii) is skipped if the cointegrating vector $a$ is specified a priori.
- The null hypothesis $H_0$ in the Engle-Granger test is no cointegration, and the alternative is cointegration.
- There are two cases:
  (i) The proposed cointegrating vector $a$ is specified *a priori*, for example, by virtue of economic theory.
  (ii) The proposed cointegrating vector $a$ is estimated from the data by means of a linear regression. In this situation, several different cases may have to be considered, depending on whether the regression has no constant term, has a constant term, or has a deterministic trend.
- Tests for cointegration using a specified cointegrating vector $a$ are more powerful than tests relying on linear regression to estimate $a$.
- The Python package `statsmodels.tsa.stattools.coint` has code for testing for cointegration in the univariate case discussed here.

# Stochastic volatility

- Another type of non-stationarity is present in models with stochastic volatility to which we shall now turn.
- In the specification of an $AR(1)$ time series model, it is assumed that the disturbances $\varepsilon_t$ are i.i.d. and $N(0, \sigma^2)$ distributed.
- Generally, this assumption is invalid in financial time series, as they typically exhibit periods of elevated and diminished volatility.
- This phenomenon is referred to as *stochastic volatility* or *heteroskedasticity*.

# EWMA model

- Conceptually, the simplest volatility model is the *exponentially weighted moving average* (EWMA) model. We have encountered this type of a model earlier in this lecture in a different context.
- It is specified as follows:

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \lambda \sigma_{t-1}^2 + (1-\lambda)\varepsilon_{t-1}^2, \end{aligned} \tag{49}$$

  where $z_t \sim N(0,1)$, and where $0 < \lambda < 1$ is a parameter. Typically, $\lambda$ is close to 1, say $\lambda = 0.97$.
- The model is very easy to use. The innovations $\widehat{\varepsilon}_t$ are calculated from the observations $x_t$ of the underlying time series as $\widehat{\varepsilon}_t = x_t - x_{t-1}$. The scaling factor $\lambda$ can be found using MLE, or set to the preferred value.
- It is a good idea to "prime" the model by adding a period prior to $t = 0$ to let the model arise the impact of the initial value of the (unobserved) $\sigma$.

# ARCH models

- An *autoregressive conditional heteroskedasticity* (ARCH) model is specified as follows:

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \zeta_0 + \zeta_1 \varepsilon_{t-1}^2 + \ldots + \zeta_q \varepsilon_{t-q}^2, \end{aligned} \tag{50}$$

  where, again, $z_t \sim N(0, 1)$ are i.i.d.. A models with this specification is denoted by $ARCH(q)$.
- The lagged terms on the RHS of the second equation describe the "moving average" character of the process.
- Notice that

$$\begin{aligned} \mathrm{E}(\varepsilon_t) &= \mathrm{E}(\mathrm{E}(\varepsilon_t | \varepsilon_{0:t-1})) \\ &= \mathrm{E}(\mathrm{E}(z_t)(\zeta_0 + \zeta_1 \varepsilon_{t-1}^2 + \ldots + \zeta_q \varepsilon_{t-q}^2)^{1/2}). \\ &= 0. \end{aligned} \tag{51}$$

# ARCH models

- The volatility process given by (50) is, in general, not covariance-stationary. It is easy to see that
$$\mathrm{E}(\sigma_t^2) = \frac{\zeta_0}{1 - \zeta_1 - \ldots - \zeta_q}\,, \tag{52}$$
which is positive if $\zeta_1 + \ldots + \zeta_q < 1$. One can, in fact, show that this condition is necessary and sufficient for covariance-stationarity of the process.
- ARCH processes exhibit *leptokurtosis* (a.k.a. *fat tails*). Consider, for example, an *ARCH*(1) model:
$$\begin{aligned} \varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \zeta_0 + \zeta_1 \varepsilon_{t-1}^2, \end{aligned} \tag{53}$$
and calculate the 4-the moment of $\sigma_t$.
- Notice first that
$$\begin{aligned} \mathrm{E}(\varepsilon_t^4) &= \mathrm{E}(\sigma_t^4)\mathrm{E}(z_t^4) \\ &= 3\mathrm{E}(\sigma_t^4). \end{aligned}$$

# ARCH models

- Next, using (53) and (52) (with $q = 1$) we find that

$$\begin{aligned} \mathrm{E}(\sigma_t^4) &= \zeta_0^2 + 2\zeta_0\zeta_1 \mathrm{E}(\varepsilon_{t-1}^2) + \zeta_1^2 \mathrm{E}(\varepsilon_{t-1}^4) \\ &= \zeta_0^2 + \frac{2\zeta_0^2\zeta_1}{1-\zeta_1} + 3\zeta_1^2 \mathrm{E}(\sigma_{t-1}^4). \end{aligned}$$

- Since $\mathrm{E}(\sigma_t^4) = \mathrm{E}(\sigma_{t-1}^4)$, we infer that

$$\mathrm{E}(\sigma_t^4) = \frac{\zeta_0^2(1+\zeta_1)}{1-3\zeta_1^2}\,, \tag{54}$$

and the kurtosis is of the residuals in the $ARCH(1)$ model is equal to

$$\frac{\mathrm{E}(\varepsilon_t^4)}{\mathrm{E}(\varepsilon_t^2)^2} = 3\,\frac{1-\zeta_1^2}{1-3\zeta_1^2}\,. \tag{55}$$

# ARCH models

- Notice that the kurtosis is greater than 3 (and, in fact, it tends to infinity as $\zeta_1 \longrightarrow 1/\sqrt{3}$), and so the tails of the distribution are heavy.
- ARCH models can be estimated using MLE. We will not discuss it here; instead we will explain the method for the more general GARCH family of models below.
- ARCH models are not frequently used in practice, because they generally lead to unrealistic volatility structures.
- They are primarily of historical interest, as they were the first stochastic volatility models proposed in the econometric literature.

# GARCH models

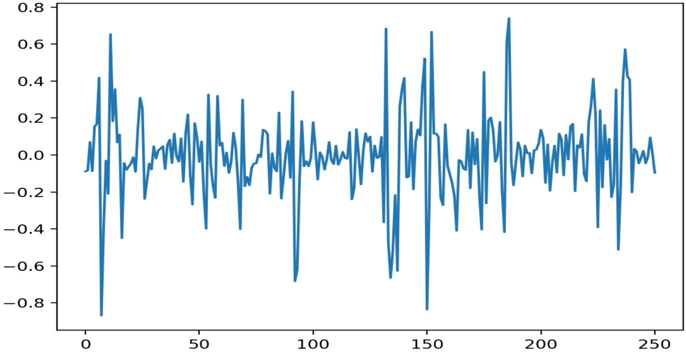
- A more general family of models is referred to as *generalized autoregressive conditional heteroskedasticity* (GARCH) models with the specification:

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \kappa + \eta_1 \sigma_{t-1}^2 + \ldots + \eta_p \sigma_{t-p}^2 + \zeta_1 \varepsilon_{t-1}^2 + \ldots + \zeta_q \varepsilon_{t-q}^2. \end{aligned} \tag{56}$$

  A model with this specification is referred to as a $GARCH(p, q)$ model.
- Compare to $ARCH(q)$, $GARCH(p, q)$ models exhibit “autoregressive” terms, in addition to the “moving average” terms, in their specifications.
- These autoregressive terms render GARCH suitable for capturing the phenomenon of *volatility clustering*, i.e. prolonged periods of elevated or lower volatility.

# GARCH models

- The graph below shows a simulated $GARCH(1,1)$ time series with the following choice of parameters: $\kappa = 0.0$, $\eta = 0.7$, $\zeta = 0.1$.



# GARCH models

- Arguments similar to the ones in the case of ARCH models show that

$$\mathrm{E}(\varepsilon_t) = 0, \tag{57}$$

and

$$\mathrm{E}(\sigma_t^2) = \frac{\kappa}{1 - \sum_{i=1}^{\min(p,q)}(\eta_i + \zeta_i)} . \tag{58}$$

It is thus necessary for stationarity of the model that $\sum_{i=1}^{\min(p,q)}(\eta_i + \zeta_i) < 1$.

- Commonly used in finance is the *GARCH*$(1, 1)$ model with

$$\sigma_t^2 = \kappa + \eta\sigma_{t-1}^2 + \zeta\varepsilon_{t-1}^2 . \tag{59}$$

- For this model, $\mathrm{E}(\sigma_t^2) = \sigma^2$ is independent of $t$, and

$$\sigma^2 = \frac{\kappa}{1 - \eta - \zeta} . \tag{60}$$

# GARCH models

- Repeating the arguments leading to formula (55), we find that the kurtosis in the *GARCH*(1, 1) model is given by

$$\frac{\mathrm{E}(\varepsilon_t^4)}{\mathrm{E}(\varepsilon_t^2)^2} = 3\,\frac{1-(\zeta+\eta)^2}{1-(\zeta+\eta)^2-2\zeta^2}\,. \tag{61}$$

- The kurtosis is always greater than 3, and thus the model exhibits fat tails. Notice also that the fourth moment exists only if $(\zeta+\eta)^2 + 2\zeta^2 < 1$.
- As an example, consider a time series of observations $x_t$ with

$$\begin{aligned} X_t &= \alpha + \beta X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \kappa + \eta\sigma_{t-1}^2 + \zeta\varepsilon_{t-1}^2. \end{aligned} \tag{62}$$

- This is an *AR*(1) model with *GARCH*(1, 1)-style residuals.

# MLE for GARCH

- It is straightforward to write the likelihood function for a model with GARCH-style disturbances. As an example, consider the model introduced in (62).
- Consider a series of observations $x_0, x_1, \ldots, x_T$. Notice that the volatility parameters $\sigma_0, \sigma_1, \ldots, \sigma_T$ are not observed data.
- The implied i.i.d. normalized shocks are given by

$$\widehat{z}_t = \frac{x_t - \alpha - \beta x_{t-1}}{\sigma_t} . \tag{63}$$

- We adopt the conditional approach, in which the likelihood function is conditioned on $x_0$ and $\sigma_0 = 0$. The joint probability distribution function can be written as

$$\begin{aligned} f(x_{1:T}|x_0, \sigma_0, \theta) &= f(x_T|x_{0:T-1}, \sigma_0, \theta) f(x_{T-1}|x_{0:T-2}, \sigma_0, \theta) \ldots f(x_1|x_0, \sigma_0, \theta) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\Big(-\frac{\widehat{z}_t^2}{2}\Big). \end{aligned}$$

# MLE for GARCH

- The volatilities $\sigma_t$ are calculated recursively from (62):

$$\begin{aligned}\sigma_1^2 &= \kappa,\\ \sigma_2^2 &= \kappa + \eta\kappa + \zeta(x_1 - \alpha - \beta x_0)^2,\\ &\dots\\ \sigma_t^2 &= \kappa + \eta\sigma_{t-1}^2 + \zeta(x_{t-1} - \alpha - \beta x_{t-2})^2.\end{aligned}$$

- This leads to the following conditional LLF:

$$-\log \mathcal{L}(\theta|x_{0:T}, \sigma_0) = \frac{1}{2}\sum_{t=1}^{T}\Big(\log \sigma_t^2 + \frac{(x_t - \alpha - \beta x_{t-1})^2}{\sigma_t^2}\Big) + const, \qquad (64)$$

where $\theta$ denotes the collection of model parameters $(\alpha, \beta, \kappa, \eta, \zeta)$. Its minimum cannot be calculate in closed form but is straightforward to calculate numerically.

# Forecasting volatility with GARCH

- A one-period volatility forecast is given by

$$\begin{aligned}\sigma^{2*}_{t+1|1:t} &= \mathsf{E}_t(\sigma^2_{t+1})\\ &= \kappa + \eta\sigma^2_t + \zeta\sigma^2_t \\ &= \sigma^2 + (\eta+\zeta)(\sigma^2_t - \sigma^2),\end{aligned} \tag{65}$$

  where $\sigma^2$ is the expected variance given by (60).
- Continuing this process we find that the $k$-step forecast is given by

$$\sigma^{2*}_{t+k|1:t} = \sigma^2 + (\eta+\zeta)^k(\sigma^2_t - \sigma^2). \tag{66}$$

- From stationarity, $\eta + \zeta < 1$, and so in the limit of long forecasting horizon,

$$\sigma^{2*}_{t+k|1:t} \longrightarrow \sigma^2, \tag{67}$$

  i.e. the forecast approaches the equilibrium value exponentially fast.

# Limitations of GARCH models

- GARCH is a stationary model. In reality, observed volatilities on financial instruments are often non-stationary.
- Stationarity requires that the drift and all $\eta$'s and $\zeta$'s have to be positive, which limits the ways in which the disturbances impact the time evolution of volatility.
- The impact of the shocks $\varepsilon_t$ on the volatility process is symmetric: positive and negative shocks have the same impact.
- This is not the case in the equity markets, where negative shocks tend to elevate volatility, while positive shocks lower it.

# Extensions of the GARCH model

- Because of the limitations of the GARCH model, various (literally, hundreds of them) extensions have been proposed. Here are a few of them.
- An *integrated model IGARCH*(1, 1) is a unit root version of *GARCH*(1, 1) with $\eta + \zeta = 1$:

$$\sigma_t^2 = \kappa + \eta\sigma_{t-1}^2 + (1-\eta)\varepsilon_{t-1}^2. \tag{68}$$

- More generally, *IGARCH*($p, q$) is specified by (56) with $\sum_{j=1}^{p} \eta_j + \sum_{j=1}^{q} \zeta_j = 1$.

# Extensions of the GARCH model

- The *exponential GARCH*$(p, q)$ *model* or *EGARCH*$(p, q)$ is defined by:

$$\log \sigma_t^2 = \kappa + \sum_{j=1}^{p} \eta_j \log \sigma_{t-1}^2 + \sum_{j=1}^{q} \zeta_j g(Z_{t-j}), \tag{69}$$

  where $g(Z) = \theta Z + \lambda(|Z| - \mathrm{E}(|Z|))$, and where $Z \sim N(0, 1)$.
- This choice of the structure of disturbances allows for asymmetric impact of $Z$ on the volatility discussed above.

# References

[1] Hamilton, J. D.: *Time Series Analysis*, Princeton University Press (1994).

[2] Tsay, R. S.: *Analysis of Financial Time Series*, Wiley (2010).