Financial Data Analysis

Basic Time Series Concepts

Summer 2014

July 15, 2014
Time Series

• A time series is a sequence of random variables in time order,

$$\{Y_t\}_{t=-\infty}^{\infty} = \{\ldots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \ldots\}.$$ 

• The index $t$ of $Y_t$ refers to time.

• We will typically just refer to “time series $Y_t$” rather than $\{Y_t\}_{t\in\mathbb{Z}}$. 
Stationarity

- *Stationarity* is a property that guarantees that essential properties of a time series remain constant over time.

- That is, time series $Y_t$ is stationary if vector

$$ (Y_1, \ldots, Y_n)' $$

has the same statistical properties as the time–shifted vector

$$ (Y_{1+h}, \ldots, Y_{n+h})' $$

for any integers $h$ and $n > 0$.

- Features that do not change over time can be exploited to systematically model and predict a time series.

- There are two forms of stationarity: *Strict* and *weak* stationarity.
Stationarity

• Roughly, strict stationarity means that all statistical properties are time-invariant, whereas weak stationarity requires the same for the second-moment structure.
Strict Stationarity

- Strict stationarity requires that the entire probabilistic structure remains constant over time, i.e., vector

\[(Y_1, \ldots, Y_n)'

has the same joint distribution function as the time–shifted vector

\[(Y_{1+h}, \ldots, Y_{n+h})'

for all integers \(h\) and \(n > 0\).

- That is, all the multivariate distributions for subsets of variables are invariant under time shift.
Strict Stationarity

- For example, with $n = 1$, this implies that the distribution function (cdf)

$$F_t(y) = \Pr(Y_t \leq y)$$

is the same for all $t$, i.e.,

$$F_t(y) = F_s(y) := F(y) \quad \text{for all } s \text{ and } t.$$  

- That is, the marginal (or unconditional) distribution of $Y_t$ does not vary over time; it is the same for all $t$. 

Weak Stationarity

- **Weak stationarity** is also known as **covariance stationarity**, or **wide-sense stationarity**, or **second-order stationarity**.

- This imposes conditions on the first two moments of the series.

- I.e., a weakly stationary series is one where the vectors $(Y_1, \ldots, Y_n)'$ and $(Y_{1+h}, \ldots, Y_{n+h})'$ have the same mean vectors and covariance matrices for all $h$ and $n > 0$. 
Weak Stationarity

- That is, time series $Y_t$ is weakly stationary if
  
  (i) the second moments are finite,

  \[ E(Y_t^2) < \infty \text{ for all } t, \]  

  (3)

  (ii) the mean function does not depend on time,

  \[ \mu_t := E(Y_t) = \mu \text{ for all } t, \]  

  (4)

  and

  (iii) the autocovariance function,

  \[ \gamma(s, t) = \text{Cov}(Y_s, Y_t) = E(Y_s Y_t) - E(Y_s) E(Y_t), \]  

  (5)

  depends only on the distance in time, $\tau = |s - t|$, between the two random variables.$^1$

\[ |E(Y_t Y_{t-\tau})| \leq \sqrt{E(Y_t^2) E(Y_{t-\tau}^2)}, \] so that (i) already implies that the covariance function exists finite.
Weak Stationarity

- Note that, for a covariance stationary process,

\[ \gamma(0) = \text{Var}(Y_t) < \infty, \]  \hspace{1cm} (6)

so conditions (i)-(iii) above imply that the variance is also time-invariant (does not depend on \( t \)).
Weak Stationarity

• For a weakly stationary process, we can define the autocovariance function at lag $\tau \in \mathbb{Z}$, $\gamma(\tau)$,

$$
\gamma(\tau) := \text{Cov}(Y_t, Y_{t-\tau})
\quad = \quad \mathbb{E}(Y_t Y_{t-\tau}) - \mathbb{E}(Y_t) \mathbb{E}(Y_{t-\tau})
\quad = \quad \mathbb{E}(Y_t Y_{t-\tau}) - \mathbb{E}^2(Y_t),
$$

and the autocorrelation function (ACF) at lag $\tau$

$$
\rho(\tau) := \text{Corr}(Y_t, Y_{t-\tau})
\quad = \quad \frac{\text{Cov}(Y_t, Y_{t-\tau})}{\sqrt{\text{Var}(Y_t)}} = \frac{\mathbb{E}(Y_t Y_{t-\tau}) - \mathbb{E}^2(Y_t)}{\sqrt{\mathbb{E}(Y_t^2) - \mathbb{E}^2(Y_t)}}
\quad = \quad \frac{\gamma(\tau)}{\gamma(0)}.
$$
Weak Stationarity

- Clearly $\gamma(\tau) = \gamma(-\tau)$. Thus, it suffices to look at the ACF for lags $\tau \geq 0$.

- The ACF provides an indication of the temporal dependence of the series and thus the extent to which it is possible to forecast a series from its own past, or the strength of the “memory” in the series.
Weak and Strict Stationarity

- The terminology suggests that strict stationarity implies weak stationarity.

- This is only the case, however, if $\mathbb{E}(Y_t^2) < \infty$.

- To see this, note that strict stationarity implies (by definition) that

\[
(Y_1, Y_{1+\tau})' \overset{d}{=} (Y_t, Y_{t+\tau})' \quad \text{for all integers } t \text{ and } \tau, \tag{7}
\]

where notation $\overset{d}{=}$ means that the two random vectors have the same distribution.
Weak and Strict Stationarity

• If \( E(Y_t^2) < \infty \), then (7) clearly implies that both \( E(Y_t) \) and \( E(Y_t^2) \) do not depend on \( t \), and

\[
\text{Cov}(Y_1, Y_{1+\tau}) = \text{Cov}(Y_t, Y_{t+\tau}) = \gamma(\tau)
\] (8)

independent of \( t \).

• Thus a strictly stationary series with finite second moment is also weakly stationary.
 Weak and Strict Stationarity

• In general, strict stationarity need not imply weak stationarity due to the requirement of the latter that the first two moments are finite.

• E.g., any independent and identically distributed (iid) series is strictly stationary.

• Thus suppose, e.g., that

$$Y_t \overset{iid}{\sim} S_\alpha, \quad 0 < \alpha < 2,$$

(9)

a non–Gaussian iid sequence of stable random variables.

• Then \( \{Y_t\} \) is strictly but not weakly stationary since the variance is not finite.
Example: White Noise

- A (weakly) stationary time series \( \{ \epsilon_t \} \) is white noise if

$$
\mu = 0, \quad \text{and} \quad 
\gamma(\tau) = \begin{cases} 
\sigma^2 & \text{for } \tau = 0 \\
0 & \text{for } \tau \neq 0,
\end{cases}
$$

that is, it is an uncorrelated zero–mean process.

- If \( \{ \epsilon_t \} \) is independent and identically distributed (iid), then it is strict (or strong, or independent) white noise.

- If, furthermore, \( \epsilon_t \overset{iid}{\sim} N(0, \sigma^2) \), then \( \{ \epsilon_t \} \) is Gaussian white noise.

- Strict white noise (as any iid series) is strictly stationary.
Example: Weak and Strict White Noise

- Consider the process
  \[ \epsilon_t = \eta_t \eta_{t-1}, \]
  where \( \eta_t \overset{iid}{\sim} \mathcal{N}(0, 1) \), i.e., \( \{\eta_t\} \) is standard normal white noise.

- Show that \( \epsilon_t \) is white noise but not strict white noise.\(^2\)

\(^2\) (G)ARCH processes are another (and practically more relevant) example of weak but not strict white noise processes.
Example: First–order moving average, or MA(1), process

- Consider the time series $Y_t$ defined by

$$Y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}, \quad t \in \mathbb{Z}, \quad (11)$$

where $\{\epsilon_t\}$ is white noise with variance $\sigma^2$, and $\mu, \theta \in \mathbb{R}$.

- Show that $E(Y_t) = \mu$, $\text{Var}(Y_t) = \sigma^2(1 + \theta^2)$, and

$$\rho(\tau) = \begin{cases} 
1, & \tau = 0 \\
\frac{\theta}{1+\theta^2}, & \tau = 1 \\
0, & \tau > 1.
\end{cases} \quad (12)$$

- Note that $\{Y_t\}$ is (weakly) stationary.
• Note that

\[ \rho(1; \theta) = \frac{\theta}{1 + \theta^2} = \frac{\theta^{-1}}{1 + \theta^{-2}} = \rho(1; \theta^{-1}). \]  

(13)
Example continued

• For the MA(1), we have $|\rho(1)| \leq \frac{1}{2}$.

• More generally, it can be shown that

$$\rho(\tau) = \begin{cases} 
1, & \tau = 0 \\
\rho, & \tau = 1 \\
0, & \tau > 1,
\end{cases} \quad (14)$$

is the ACF of a stationary time series if and only if $|\rho| \leq \frac{1}{2}$.

• This can be seen by showing that with $|\rho| > \frac{1}{2}$ the ACF (14) is not nonnegative definite.
Example continued

• If time series \( \{Y_t, t \in \mathbb{Z}\} \) existed with unit variance\(^3\) and ACF (14) with \(|\rho| > \frac{1}{2}\), then, with constants \(a_i, i = 1, \ldots, n\),

\[
\text{Var} \left( \sum_{i=1}^{n} a_i Y_i \right) = \sum_{i=1}^{n} a_i^2 \text{Var}(Y_i) + 2 \sum_{i=2}^{n} \sum_{j<i} a_i a_j \text{Cov}(Y_i, Y_j)
\]

\[
= n + 2\rho \sum_{i=2}^{n} a_i a_{i-1}.
\]

• With \(a_1 = 1, a_2 = -1, a_3 = 1, \ldots, a_n = (-1)^{n+1}\), this becomes

\[
\text{Var} \left( \sum_{i=1}^{n} a_i Y_i \right) = n - 2\rho(n - 1) < 0 \quad \text{for } n > \frac{2\rho}{2\rho - 1}. \quad (15)
\]

\(^3\)Assumed for simplicity.
Example continued

- Thus there is no 1–correlated time series with $\rho = \rho(1) > \frac{1}{2}$.\(^4\)

- For $\rho < -\frac{1}{2}$, the same argument can be repeated with $a_1 = a_2 = \cdots = a_n = 1$.

- Find an intuitive explanation for this fact.

---

\(^4\)A stationary time series is referred to as $q$–correlated if $\rho(\tau) = 0$ for $\tau > q$. 
Sample AutoCorrelation Function (SACF)

- In practice, before specifying a statistical model, we will look at the data.

- To assess the degree of time–dependence in the data, and to select a suitable model that reflects these properties, an important tool is the Sample AutoCorrelation Function (SACF).

- With a sample of length $T$, $Y_1, Y_2, \ldots, Y_T$, let

$$
\hat{\mu} = \bar{Y} = \frac{1}{T} \sum_{t=1}^{T} Y_t
$$

$$
\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} (Y_{t+\tau} - \bar{Y})(Y_t - \bar{Y}) \quad (16)
$$

$$
\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} \quad (SACF \ at \ lag \ \tau).
$$
Sample AutoCorrelation Function (SACF)

- Note that in (16), only \( T - \tau \) terms enter the summation, yet the ACF is typically estimated by (16) rather than by

\[
\hat{\gamma}(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} (Y_{t+\tau} - \bar{Y}_T)(Y_t - \bar{Y}_T). \tag{17}
\]

- The reason is that, in contrast to (17), use of (16) guarantees that the resulting estimate of the covariance matrix of \( m \) consecutive observations is positive (semi-)definite,\(^5\) i.e., matrices of the form

\[
\hat{\Gamma}_m = \begin{pmatrix}
\hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(m-1) \\
\hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(m-2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\gamma}(m-1) & \hat{\gamma}(m-2) & \cdots & \hat{\gamma}(0)
\end{pmatrix}.
\]

\(^5\) \( \hat{\Gamma}_m \) is positive definite if \( \hat{\gamma}(0) > 0 \), cf. Brockwell and Davis, 2002. *Introduction to Time Series and Forecasting*, Springer, p. 59.
Sample AutoCorrelation Function (SACF)

- Both (16) and (17) are biased, but\(^6\) nearly unbiased for large sample sizes (and small \(\tau\), relative to \(T\)).

- For large \(\tau\), relative to \(T\), the SACF \(\hat{\rho}(\tau)\) is an unreliable estimator of \(\rho(\tau)\) due to the scarcity of pairs \((Y_t, Y_{t+\tau})\) (e.g., there is only one pair if \(\tau = T - 1\)).

- As a rule of thumb, one would only look at \(\hat{\rho}(\tau)\) when \(T \geq 50\) and for \(\tau \leq T/4\).

- A plot of the SACF is referred to as sample correlogram.

---

\(^6\)Subject to some conditions.
Sample AutoCorrelation Function (SACF)

- It is often of interest to test whether a given series is white noise.

- One approach is to test the hypothesis that the autocorrelation coefficients are all zero.

- To do so, the sampling distribution of the SACF is required.
Sample AutoCorrelation Function (SACF)

• If the process under study is strict (independent) white noise, then (subject to a further moment condition)\(^7\)

\[
\sqrt{T} \hat{\rho}(\tau) \xrightarrow{d} N(0, 1), \quad \tau > 0,
\]  

(18)

and the sample autocorrelations at different lags \(\tau\) are asymptotically independent.

• Thus, under the null hypothesis of strict white noise, we can treat \(\hat{\rho}(\tau)\) as asymptotically normal with variance \(1/T\), i.e.,

\[
\hat{\rho}(\tau) \xrightarrow{a} N\left(0, \frac{1}{T}\right).
\]  

(19)

Sample AutoCorrelation Function (SACF)

- Based on result (18), $100 \times (1 - \alpha)\%$ asymptotic confidence intervals, given by
  $$\pm z_{\alpha/2} \frac{1}{\sqrt{T}},$$
  are often added to the sample correlogram, where $z_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution.

- On the next slide, dashed lines represent 95% asymptotic confidence intervals associated with a strict white noise process ($\alpha = 0.5$, $z_{\alpha/2} = -1.96 \approx 2$).
Sample AutoCorrelation Function (SACF)

autocorrelations of S&P 500 returns

autocorrelations of DAX 30 returns
SACF: Box–Pierce and Ljung–Box statistics

- Even if all population autocorrelations are zero, some individual sample autocorrelations will fall outside a 95% confidence interval due to random noise (one expects one in 20).

- Thus, rather than testing the autocorrelation coefficients one–by–one, we might want to test the joint hypothesis

  \[ H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0 \]  

  against the alternative \( H_1 : \rho(\ell) \neq 0 \) for some \( \ell \in \{1, \ldots, m\} \).

- Since the square of a standard normal is \( \chi^2(1) \), and the sum of independent \( \chi^2 \)s is likewise \( \chi^2 \) with the degrees of freedom adding up, the Box–Pierce statistic

  \[ Q(m) = T \sum_{\tau=1}^{m} \hat{\rho}^2(\tau) \xrightarrow{d} \chi^2(m) \]  

  can be used to test (20).
SACF: Box–Pierce and Ljung–Box statistics

• A finite–sample correction to (21) with faster convergence to the $\chi^2(m)$
distribution was suggested by Ljung and Box (1978), namely

$$\tilde{Q}(m) = T(T + 2) \sum_{\tau=1}^{m} \frac{\hat{\rho}^2(\tau)}{T - \tau}, \quad (22)$$

which has the same limiting distribution as $Q(m)$ in (21), i.e., $\chi^2(m)$.

• Large values of $\tilde{Q}(m)$ (or $Q(m)$) provide evidence against $H_0$ in (20).

• The null hypothesis (20) is rejected at level $\alpha$ if

$$\tilde{Q}(m) > F^{-1}_{\chi^2(m)}(1 - \alpha), \quad (23)$$

where $F^{-1}_{\chi^2(m)}(1 - \alpha)$, is the $1 - \alpha$–quantile of the $\chi^2$ distribution with $m$ degrees of freedom.
SACF: Box–Pierce and Ljung–Box statistics

• How to select \( m \)?

• Choosing \( m \) too small may result in significant higher–order autocorrelations being missed.

• Selecting \( m \) too large, on the other hand, may reduce the power of the test due to the dominance of insignificant higher–order coefficients.

• Thus, in practice, \( \tilde{Q}(m) \) is often reported for several values of \( m \).
Sample AutoCorrelation Function (SACF)

- Confidence intervals/teststatistics based on this theory are routinely computed (and plotted along with sample correlograms).
- However, it should be kept in mind that the results require independent white noise.
- If the process is uncorrelated but not independent (e.g. GARCH), then \( \hat{\rho}(\tau) \) typically has a higher asymptotic variance.
- Thus, when the test statistics described above exceed the respective critical values, we can only reject the null of independent white noise, whereas they do not allow to test just for uncorrelatedness.
- See, e.g., Taylor (2005), p. 48 and p. 80, for further discussion.
Lag Operators

• The lag operator, denoted by $L$, is an operator that shifts the time index backward by one unit.

• Applying the lag operator to a variable at time $t$, we obtain the variable at time $t-1$, i.e.,

$$LY_t = Y_{t-1}.$$

• Applying $L^2 = L \cdot L$ or $L^3$ amounts to lagging the variable twice or thrice (etc.), i.e.,

$$L^2Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}$$

in general: $L^qY_t = Y_{t-q}.$
Lag Operators

• A constant $c$ may be viewed as a special series $\{Y_t\}_{t \in \mathbb{Z}}$ with $Y_t = c$ for all $t$; hence

$$Lc = c.$$  \hfill (24)

• Polynomials in $L$ can be handled just as polynomials in “regular” variables, e.g.,

$$(1 - L)^2 Y_t = (1 - L)(1 - L)Y_t$$ \hfill (25)
$$= (1 - 2L + L^2)Y_t$$ \hfill (26)
$$= Y_t - 2Y_{t-1} + Y_{t-2}$$ \hfill (27)
$$= \Delta^2 Y_t = \Delta(\Delta Y_t),$$ \hfill (28)

where the lag–1 difference operator

$$\Delta = 1 - L, \quad \Delta Y_t = Y_t - Y_{t-1}.$$ \hfill (29)
Lag Operators

- E.g., if $P_t$ is the price of an asset, then the log–return

$$r_t = \log P_t - \log P_{t-1} = \Delta \log P_t.$$  \hspace{1cm} (30)
ARMA Models

- A *time series model* is a specification of the joint distributions (or only the second–moment structure) of a sequence of random variables.

- **AutoRegressive Moving Average** time series models are often used to model dynamics in the *conditional mean* of a time series.

- The techniques used to study these processes are also useful for financial time series models such as ARCH/GARCH.
Moving Average (MA) Processes

- A moving average process of order $q$, or MA($q$) process, is defined by

$$Y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \quad \theta_q \neq 0,$$

(31)

$$= \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \epsilon_t$$

$$= \mu + \theta(L) \epsilon_t,$$

where $\{\epsilon_t\}$ is a white noise process with variance $\sigma^2$, and the lag polynomial

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q,$$

(32)

with the lag operator $L \ (L^m x_t = x_{t-m})$.

- $Y_t$ is a weighted **average** of a white noise process which **moves** through time.
Moving Average (MA) Processes

- The moments of the MA(1) process have already been calculated above.

- Observe that
  
  (i) $Y_t$ is weakly stationary (for any value of $\theta$).
  
  (ii) If $\epsilon_t$ is normal, i.e., $\epsilon_t \sim \text{iid} N(0, \sigma^2)$, then the marginal distribution of the process is likewise normal,

  $$Y_t \sim N(\mu, \sigma^2(1 + \theta^2)),$$

  for all $t$,

  and, more generally, any subset of variables is (multivariate) normal,\(^8\) i.e., for any $t$ and $h$,

  $$
  \begin{bmatrix}
  y_t \\
  y_{t-1} \\
  y_{t-2} \\
  \vdots \\
  y_{t-h}
  \end{bmatrix}
  \sim
  N
  \left(\begin{bmatrix}
  \mu \\
  \mu \\
  \mu \\
  \vdots \\
  \mu
  \end{bmatrix},
  \sigma^2
  \begin{bmatrix}
  1 + \theta^2 & \theta & 0 & \cdots & 0 \\
  \theta & 1 + \theta^2 & \theta & \cdots & 0 \\
  0 & \cdots & \theta & 1 + \theta^2 & \theta \\
  \vdots & \cdots & \vdots & \theta & 1 + \theta^2 \\
  0 & \cdots & 0 & \theta & 1 + \theta^2
  \end{bmatrix}
  \right)\)
  $$

\(^8\)Clearly $\{Y_t\}$ is then also strictly stationary.
Moving Average (MA) Processes

- These results generalize to MA($q$) processes:

(i) Straightforward calculation shows that, for the MA($q$),

$$Y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t,$$

(33)

$$E(Y_t) = \mu,$$ the variance is, with $\theta_0 = 1$,

$$\gamma(0) = \sigma^2 \sum_{i=0}^{q} \theta_i^2,$$

and the autocorrelation function (ACF)

$$\rho(\tau) = \begin{cases} \frac{\sum_{i=0}^{q-\tau} \theta_i \theta_{i+\tau}}{\sum_{i=0}^{q} \theta_i^2} & \text{for } \tau \leq q \\ 0 & \text{for } \tau > q. \end{cases}$$

(34)

We observe that the ACF of an MA($q$) process cuts off (equals zero) after lag $q$.$^9$

$^9$That is, the MA($q$) process is $q$–correlated.
(ii) Finite–order \((q < \infty)\) MA\(q)\) processes are weakly stationary, since the second moments are finite and do not depend on \(t\).

(iii) If \(\{\epsilon_t\}\) is Gaussian white noise, then the distribution of any subset of variables \((Y_t, Y_{t+1}, \ldots, Y_{t+h})'\) is also (multivariate) normal.
MA(∞) Processes

Consider the MA(∞) process given by

\[ Y_t = \mu + \epsilon_t + \sum_{i=1}^{\infty} \theta_i \epsilon_{t-i}, \]

(35)

where \( \{\epsilon_t\} \) is white noise and

\[ \sum_{i=1}^{\infty} |\theta_i| < \infty, \]

(36)

i.e., the sequence of MA coefficients, \( \theta_i, i = 1, 2, \ldots \), is absolutely summable.
MA($\infty$) Processes

- Process (35) and (36) is weakly stationary with (defining $\theta_0 = 1$)

\[
E(Y_t) = \mu, \tag{37}
\]

\[
\gamma(0) = \sigma^2 \sum_{i=0}^{\infty} \theta_i^2, \tag{38}
\]

\[
\gamma(\tau) = E(Y_t Y_{t-\tau}) - \mu^2 \tag{39}
\]

\[
= E \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_i \theta_j \epsilon_t \epsilon_{t-i} \epsilon_{t-\tau-j} \right)
\]

\[
= \sum_i \sum_j \theta_i \theta_j E(\epsilon_t \epsilon_{t-i} \epsilon_{t-\tau-j})
\]

\[
= \sigma^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}, \tag{40}
\]

\[
\rho(\tau) = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}}{\sum_{i=0}^{\infty} \theta_i^2}. \tag{41}
\]
MA($\infty$) Processes

- **MA($\infty$):**
  \[
  Y_t = \mu + \varepsilon_t + \sum_{i=1}^{\infty} \theta_i \varepsilon_{t-i}, \quad \sum_{i=1}^{\infty} |\theta_i| < \infty, \tag{42}
  \]
  and \(\{\varepsilon_t\}\) is white noise.

- If \(\{\varepsilon_t\}\) is strict (independent) white noise, then process (42) is said to be a **linear time series**.

- As for finite-order MA processes, if \(\{\varepsilon_t\}\) is Gaussian white noise, then the distribution of any subset of variables is also (multivariate) normal.

- A time series (e.g., a GARCH process) may be a linear function of (non–independent) white noise but a **nonlinear** function of strict white noise; such a process belongs to the class of nonlinear processes.
Autoregressive (AR) Processes

• An autoregressive process of order \( p \), abbreviated AR\((p)\), is of the form

\[
Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \quad \phi_p \neq 0,
\]

\[
\phi(L)Y_t = \epsilon_t,
\]

where \( \epsilon_t \) is white noise with mean zero and variance \( \sigma^2 \), and lag polynomial

\[
\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p,
\]

with \( L \) being the lag operator, \( L^m x_t = x_{t-m} \).

• The value of the time series at time \( t \) is a linear regression on its own past.
**AR(1) Process**

- Consider the first–order AR process,

\[
Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(\sigma^2),
\]

where “\(\text{WN}(\sigma^2)\)” means white noise with variance \(\sigma^2\).

- Now assuming that

(i) \(\{Y_t\}\) is stationary,
(ii) \(\epsilon_t\) is uncorrelated with \(Y_s\) for \(s < t\),\(^{10}\)

we get, for the mean,

\[
E(Y_t) = c + \phi E(Y_{t-1}) + E(\epsilon_t)
\]

\[
E(Y_t) = \frac{c}{1 - \phi},
\]

since \(E(Y_t) = E(Y_{t-1})\) by stationarity, and \(E(\epsilon_t) = 0\).

\(^{10}\)In this case \(\{Y_t\}\) is referred to as *causal*; noncausal processes will not be considered.
• For the variance, 

\[
\text{Var}(Y_t) = \varphi^2 \text{Var}(Y_{t-1}) + \text{Var}(\epsilon_t) + 2\varphi \text{Cov}(Y_{t-1}\epsilon_t) \]

\[= \varphi^2 \text{Var}(Y_t) + \sigma^2 + 0,\]

that is

\[
\text{Var}(Y_t) = \frac{\sigma^2}{1 - \varphi^2}, \tag{47}
\]

so that

\[|\varphi| < 1 \tag{48}\]

is required in this case.

• \(|\varphi| < 1\) is the stationarity condition for the AR(1) process.
**AR(1) Process**

- For the correlation structure, assuming $c = 0$,\(^{11}\) multiply (45) by $Y_{t-\tau}$, $\tau \geq 1$, and take expectations,

\[
E(Y_t Y_{t-\tau}) = \phi E(Y_{t-1} Y_{t-\tau}) + E(\epsilon_t Y_{t-\tau})
\]

\[
\gamma(\tau) = \phi \gamma(\tau - 1), \quad \tau > 0
\]

\[
\gamma(\tau) = \phi^\tau \gamma(0) = \frac{\phi^\tau \sigma^2}{1 - \phi^2}, \quad \tau \geq 0,
\]

and the autocorrelation function (ACF) at lag $\tau$,

\[
\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \phi^\tau. \quad (49)
\]

\(^{11}\)With $c = 0$, we have $E(Y_t) = 0$ and thus $\text{Cov}(Y_t Y_{t-\tau}) = E(Y_t Y_{t-\tau}) - E^2(Y_t) = E(Y_t Y_{t-\tau})$, and computations simplify a bit. Since adding a constant does not affect the correlation structure, exactly the same result is obtained with $c \neq 0$. 


AR(1) Process

- $\phi$ determines the (geometric) decay of the ACF.
AR(1) Process

- Note that the ACF of the AR process decays to zero (exponentially) and does not cut off (become zero) after a finite lag, as for moving average processes.
AR(1) Process

- Consider an AR(1) process initialized at time 0 with mean $\mu_0$ and variance $\sigma_0^2$, i.e.,

$$Y_0 \sim (\mu_0, \sigma_0^2)$$

$$(50)$$

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad t \geq 1, \quad |\phi| < 1,$$

$$(51)$$

$$\epsilon_t \sim \text{WN}(\sigma^2),$$

$$(52)$$

and $Y_0$ and $\{\epsilon_t\}$ are uncorrelated.
AR(1) Process

- Solving (50)–(52),

\[ Y_t = \phi Y_{t-1} + \epsilon_t \]
\[ = \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \]
\[ = \phi^2 Y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \]
\[ = \phi^3 Y_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \]
\[ \vdots \]
\[ = \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i \epsilon_{t-i}. \]  

(53)
AR(1) Process

- The mean of $Y_t$ is

$$E(Y_t) = \phi^t \mu_0,$$

and the variance

$$\text{Var}(Y_t) = \text{Var} \left( \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i \epsilon_{t-i} \right)$$

$$= \phi^{2t} \sigma^2_0 + \sum_{i=0}^{t-1} \phi^{2i} \text{Var}(\epsilon_{t-i})$$

$$= \frac{\sigma^2}{1 - \phi^2} + \phi^{2t} \left( \sigma^2_0 - \frac{\sigma^2}{1 - \phi^2} \right).$$

- (Also show that $\text{Corr}(Y_t, Y_{t+\tau}) = \phi^\tau \sqrt{\text{Var}(Y_t)/\text{Var}(Y_{t+\tau})}$ for $\tau > 0$.)
AR(1) Process

• Thus \( \{Y_t\} \) is (weakly) stationary if \( \mu_0 = 0 \) and \( \sigma_0^2 = \frac{\sigma^2}{1 - \phi^2} \) (process initialized with stationary moments).

• In any case, however, since \( |\phi| < 1 \) has been assumed, as \( t \) becomes large,

\[
\begin{align*}
E(Y_t) & \approx 0 \quad (56) \\
\text{Var}(Y_t) & \approx \frac{\sigma^2}{1 - \phi^2} \quad (57) \\
\text{Corr}(Y_t, Y_{t-\tau}) & \approx \phi^\tau, \quad (58)
\end{align*}
\]

i.e., the impact of initialization wears off and the process is asymptotically stationary (i.e., converges to stationarity from any initial distribution with finite variance).

• The process is also stationary if it has been initialized in the indefinite past, since then the impact of any initial conditions has disappeared.\(^{12}\)

\(^{12}\)In general, the definition of a process is complete only if some kind of initialization has been assumed.
AR(1) Process

- As for MA processes, if the white noise is Gaussian, i.e.,

\[ \epsilon_t \overset{iid}{\sim} N(0, \sigma^2), \]  

then the distribution of \((Y_t, Y_{t+1}, \ldots, Y_{t+h})'\) is also normal, i.e., for a stationary AR(1) with (59),

\[ Y_t \sim N \left( \frac{c}{1 - \phi}, \frac{\sigma^2}{1 - \phi^2} \right), \]  

and, e.g.,

\[
\begin{bmatrix}
Y_t \\
Y_{t+\tau}
\end{bmatrix}
\sim
N
\left( \begin{bmatrix}
\mu \\
\mu
\end{bmatrix}, \begin{bmatrix}
\gamma(0) & \gamma(\tau) \\
\gamma(\tau) & \gamma(0)
\end{bmatrix} \right)
\]

\[
= N \left( \begin{bmatrix}
\mu \\
\mu
\end{bmatrix}, \frac{\sigma^2}{1 - \phi^2} \begin{bmatrix}
1 & \phi^\tau \\
\phi^\tau & 1
\end{bmatrix} \right),
\]

where \( \mu = c/(1 - \phi). \)
AR(1) Process

- For $|\phi| > 1$ the process displays explosive behavior.

- The (nonstationary) case $\phi = 1$ is the *random walk* and will be considered later.

- For various values of $\phi$, the next slides show simulated paths from the AR(1) process

\[
\begin{align*}
Y_0 &= 0 \\
Y_t &= c + \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{iid} \ N(0, 1), \quad t \geq 1.
\end{align*}
\]

- E.g., for $c = 1$ and $\phi = 0.95$, as in the upper panel on the next slide, with (63), observe that the process needs a certain burn–in period until it approaches stationarity and starts to fluctuate around its stationary mean $c/(1 - \phi) = 1/(1 - 0.95) = 20$. 
\( \phi = 0.95 \)

\( \phi = -0.95 \)
process $Y_t = 1.1 Y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$, $Y_0 \sim N(0,1)$
process $Y_t = (-1.1)Y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$, $Y_0 \sim N(0,1)$
**MA(∞) representation of AR(1)**

- Stationary AR(1):

  \[ Y_t = \phi Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \quad |\phi| < 1. \quad (64) \]

- As before, recursively substituting, for \( i = 1, \ldots, \tau - 1 \),

  \[ Y_{t-i} = \phi Y_{t-i-1} + \epsilon_{t-i}, \quad (65) \]

  into (64), we obtain

  \[ Y_t = \phi^\tau Y_{t-\tau} + \sum_{i=0}^{\tau-1} \phi^i \epsilon_{t-i}. \quad (66) \]
MA(∞) representation of AR(1)

- Since $|\phi| < 1$,
  \[ \phi^\tau Y_{t-\tau} \rightarrow 0, \quad \tau \rightarrow \infty, \]  
  and so the MA(∞) representation

  \[ Y_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}. \]  

- With strong (independent) white noise \( \{\epsilon_t\} \), this is a *linear process*. 
**MA(∞) representation of AR(1)**

Alternatively, in terms of lag operators, writing

$$(1 - \phi L)Y_t = c + \epsilon_t,$$  \hspace{1cm} (69)

we can, for $|\phi| < 1$, invert the lag polynomial $1 - \phi L$ as

$$\frac{1}{1 - \phi L} = 1 + \phi L + \phi^2 L^2 + \cdots$$  \hspace{1cm} (70)

$$= \sum_{i=0}^{\infty} \phi^i L^i,$$  \hspace{1cm} (71)

in analogy to the geometric series

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1 - x} \text{ for } |x| < 1.$$  \hspace{1cm} (72)
MA(∞) representation of AR(1)

- Doing so, we get\(^{13}\)

\[
\begin{align*}
Y_t &= c + \phi Y_{t-1} + \epsilon_t \\
Y_t - \phi Y_{t-1} &= c + \epsilon_t \\
(1 - \phi L)Y_t &= c + \epsilon_t \\
Y_t &= \frac{c}{1 - \phi L} + \frac{\epsilon_t}{1 - \phi L} \\
&= \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i L^i \epsilon_t \\
&= \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}.
\end{align*}
\]

\(^{13}\)Note that, for constant \(c\), due to (24), \((1 - \phi L)^{-1} c = c + \phi L c + \phi^2 L^2 c + \cdots = (1 + \phi + \phi^2 + \cdots) c = c/(1 - \phi)\).
For the general AR($p$) process,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$  \hspace{1cm} (74)

the stationarity condition is that the roots of the *characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p-1} z^{p-1} - \phi_p z^p = 0$$  \hspace{1cm} (75)

are larger than one in magnitude ("outside the unit circle").
Autoregressive Moving Average (ARMA) Models

- ARMA($p, q$) is given by

$$\phi(L)Y_t = c + \theta(L)\epsilon_t,$$  \hspace{1cm} (76)

where the lag polynomials

$$\phi(L) = 1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p$$ \hspace{1cm} (77)

$$\theta(L) = 1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q,$$ \hspace{1cm} (78)

i.e., a combination of AR($p$) and MA($q$),

$$Y_t = c + \phi_1Y_{t-1} + \phi_2Y_{t-2} + \cdots + \phi_pY_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}.$$  

- $\phi(z)$ and $\theta(z)$ have no roots in common.
• Mixing MA and AR parts often leads to more flexible models with less parameters than using pure MA or AR models.

• As finite-order MA processes are stationary, the stationarity of an ARMA model depends on the autoregressive polynomial $\phi(z)$, and the stationarity condition is identical to that for AR($p$) processes.

• In a stationary ARMA with Gaussian white noise, all the marginal and joint distributions are likewise normal.
Example: ARMA(1,1)

- For example, consider the ARMA(1,1) process,

\[ Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t, \]  

which, if $|\phi| < 1$, can be inverted to give the MA($\infty$) representation,

\[ Y_t = \frac{1 + \theta L}{1 - \phi L} \epsilon_t = (1 + \theta L) \sum_{i=0}^{\infty} \phi^i L^i \epsilon_t \]

\[ = \epsilon_t + (\theta + \phi) \sum_{i=1}^{\infty} \phi^{i-1} \epsilon_{t-i} \]

\[ = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \]  

with $\psi_0 = 1$, $\psi_i = (\theta + \phi)\phi^{i-1}$, $i \geq 1$. 
ACF of ARMA(1,1)

- It is instructive (and will be of further use) to compare the ACF of an ARMA(1,1) with that of an AR(1) process. Thus consider

\[ Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t, \]  

(82)

we have, multiplying (82) by \( Y_t \) and taking expectations,

\[ E(Y_t^2) = \gamma(0) = \phi E(Y_{t-1}^2) + \theta E(\epsilon_{t-1} Y_t) + E(\epsilon_t Y_t), \]  

(83)

where

\[ E(\epsilon_t Y_t) = E(\epsilon_t (\phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t)) = \sigma^2, \]  

(84)

since

\[ E(\epsilon_t Y_{t-1}) = 0 \]
\[ E(\epsilon_t \epsilon_{t-1}) = 0 \]
\[ E(\epsilon_t^2) = \sigma^2. \]
ACF of ARMA(1,1)

• Moreover,

\[
E(\epsilon_{t-1}Y_t) = E(\epsilon_{t-1}(\phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t)) \\
= \phi E(\epsilon_{t-1}Y_{t-1}) + \theta E(\epsilon_{t-1}^2) + E(\epsilon_{t-1}\epsilon_t) \\
= (\phi + \theta)\sigma^2.
\]

• In summary, (83) can be written as

\[
\gamma(0) = \phi \gamma(1) + \theta(\phi + \theta)\sigma^2 + \sigma^2 \quad (85)
\]

\[
= \phi \gamma(1) + (1 + \theta(\phi + \theta))\sigma^2.
\]
ACF of ARMA(1,1)

• Similarly, multiplying (82) by $Y_{t-1}$ and taking expectations,

$$
\gamma(1) = \phi E(Y_{t-1}^2) + \theta \underbrace{E(\epsilon_{t-1}Y_{t-1})}_{\sigma^2} + \underbrace{E(\epsilon_t\epsilon_{t-1})}_{=0} \\
= \phi \gamma(0) + \theta \sigma^2,
$$

(86)

and then, multiplying (82) by $Y_{t-\tau}$, $\tau \geq 2$,

$$
\gamma(\tau) = \phi \underbrace{E(Y_{t-1}Y_{t-\tau})}_{\gamma(\tau-1)} + \theta \underbrace{E(\epsilon_{t-1}Y_{t-\tau})}_{=0} + \underbrace{E(\epsilon_tY_{t-\tau})}_{=0} \\
= \phi \gamma(\tau - 1), \quad \tau \geq 2.
$$

(87)
ACF of ARMA(1,1)

• (85) and (86), i.e.,

\[
\begin{bmatrix}
1 & -\phi_1 \\
-\phi_1 & 1
\end{bmatrix}
\begin{bmatrix}
\gamma(0) \\
\gamma(1)
\end{bmatrix} = \sigma^2
\begin{bmatrix}
1 + \theta(\phi + \theta) \\
\theta
\end{bmatrix},
\]

can be solved for the variance and the first–order autocovariance,

\[
\begin{align*}
\gamma(0) &= \frac{\sigma^2(1 + 2\phi\theta + \theta^2)}{1 - \phi^2}, \\
\gamma(1) &= \frac{\sigma^2(\theta + \phi + \theta\phi^2 + \theta^2\phi)}{1 - \phi^2} \\
&= \frac{\sigma^2(\theta + \phi)(1 + \theta\phi)}{1 - \phi^2}.
\end{align*}
\]
ACF of ARMA(1,1)

• Hence, along with (87)$^{14}$

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \phi^{\tau-1} \frac{(\phi + \theta)(1 + \phi \theta)}{1 + 2\theta \phi + \theta^2}, \quad \tau \geq 1. \quad (88)$$

$^{14}$Note that $\rho(\tau) = 0$ if $\theta = -\phi$. Then the lag polynomials $\phi(z)$ and $\theta(z)$ have a common root and $Y_t$ is white noise, i.e., in representation $\phi(L)Y_t = (1 - \phi L)Y_t = \theta(L)\epsilon_t = (1 + \theta L)\epsilon_t$, with $\theta = -\phi$, so that $(1 - \phi L)Y_t = (1 - \phi L)\epsilon_t$, canceling out the common factor $1 - \phi L$ shows that $Y_t = \epsilon_t$, i.e., white noise.
Autocorrelation function of ARMA(1,1) with $\phi = 0.8$

$$\rho(\tau) = \phi^{\tau-1} \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\theta\phi + \theta^2}$$
ACF of ARMA(1,1)

• This shows that the ACF of the ARMA(1,1) decays to zero at rate $\phi$ just as the AR(1).

• However, there is an additional degree of flexibility in the sense that the ACF may start at a low value at $\tau = 1$ but still displays a slow decay, say at rate $\phi = 0.975$.

• With the AR(1), the rate of decay and the first–order autocorrelation are always determined by the same parameter $\phi$, so level and decay of the ACF cannot be disentangled.

• Looking at the ACF of the squares of the S&P 500 returns, this anticipates that an ARMA(1,1)–type process for the squares will be more appropriate than an AR(1)–type model.
autocorrelations of absolute S&P 500 returns

autocorrelations of squared S&P 500 returns
Predictive densities and forecasting

- Let the variables observed until time $t$ be collected in the information set
  $I_t = \{ Y_s : s \leq t \}$.  

- Then the $h$–step conditional expectation

\[
E_t(Y_{t+h}) = E(Y_{t+h} | I_t) = E(Y_{t+h} | Y_t, Y_{t-1}, \ldots),
\]  \hspace{1cm} (89)

and the $h$–step conditional variance

\[
Var_t(Y_{t+h}) = Var(Y_{t+h} | I_t) = Var(Y_{t+h} | Y_t, Y_{t-1}, \ldots).
\]  \hspace{1cm} (90)

- $h$ is the forecast horizon, $t$ is the forecast origin.

---

\textsuperscript{15}In practice, other predetermined variables may enter the information set used for forecasting.
Predictive densities and forecasting

• Define the \( h \)–step–ahead predictor \( \hat{Y}_t(h) \).

• The mean–squared prediction error (MSE) of forecast \( \hat{Y}_t(h) \) is

\[
\text{MSE}(\hat{Y}_t(h)) = E[(Y_{t+h} - \hat{Y}_t(h))^2].
\]  

(91)

• We want to find the predictor which has the smallest possible MSE.

• This is given by the conditional expectation \( E_t(Y_{t+h}) = E(Y_{t+h}|I_t) \).
Predictive densities and forecasting

- For example, assume a stationary AR(1),

\[ Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad |\phi| < 1, \tag{92} \]

with \( \{\epsilon_t\} \) being strict white noise (i.e., a linear time series model).

- Then

\[ Y_{t+h} = \phi^h Y_t + \frac{c(1 - \phi^h)}{1 - \phi} + \sum_{i=0}^{h-1} \phi^i \epsilon_{t+h-i}, \]

so that the conditional expectation is\(^{16}\)

\[ E_t(Y_{t+h}) = \phi^h Y_t + \frac{c(1 - \phi^h)}{1 - \phi} = \frac{c}{1 - \phi} + \phi^h \left( Y_t - \frac{c}{1 - \phi} \right). \]

\(^{16}\)Independence of the white noise process implies \( E_t(\epsilon_{t+\tau}) = 0 \) for \( \tau > 0. \)
Predictive densities and forecasting

- In financial (and other) applications, we are not only interested in the conditional mean as a point predictor but also in the uncertainty associated with future realizations.

- Since $\text{Var}_t(\epsilon_{t+\tau}) = \text{Var}(\epsilon_{t+\tau})$ for a strict white noise process, the conditional variance is

\[
\text{Var}_t(Y_{t+h}) = \text{Var}_t \left( \sum_{i=0}^{h-1} \phi^i \epsilon_{t+h-i} \right)
\]

\[
= \sum_{i=0}^{h-1} \phi^{2i} \text{Var}_t(\epsilon_{t+h-i})
\]

\[
= \sum_{i=0}^{h-1} \phi^{2i} \text{Var}(\epsilon_{t+h-i})
\]

\[
= \sigma^2 \sum_{i=0}^{h-1} \phi^{2i}.
\]
Predictive densities and forecasting

• Note that\(^{17}\)

(i) If \(\epsilon_t \sim iid N(0, \sigma^2)\), then all predictive densities are normal, i.e.,

\[
Y_{t+h} | I_t \sim N \left( \phi^h Y_t + c \sum_{i=0}^{h-1} \phi^i, \sigma^2 \sum_{i=0}^{h-1} \phi^{2i} \right).
\]

(ii) \(E_t(Y_{t+h}) \to E(Y_t)\) as \(h \to \infty\).

(iii) \(\text{Var}_t(Y_{t+h})\) increases with \(h\) (longer forecast horizon \(\Rightarrow \) greater uncertainty) and converges to \(\text{Var}(Y_t)\) as \(h \to \infty\).

\(^{17}\)These properties, obviously with modification of the conditional mean and conditional variance in (93), apply to all linear time series models, such as general (stationary) ARMA\((p, q)\) processes with independent white noise.
(iv) In contrast to the conditional mean, the conditional variance $\text{Var}_t(Y_{t+h})$ depends only on the forecast horizon $h$ and not on $\{Y_t, Y_{t-1}, \ldots\}$, i.e., the current and past process values.

This is particularly unrealistic in financial applications, where volatility appears to be predictable based on the return history (volatility clustering).

This points to the need for nonlinear models, i.e., processes with non-iid error terms $\epsilon_t$. 
Exact and conditional maximum likelihood estimation of time series models

- To illustrate, consider a stationary Gaussian AR(1) process.

- I.e., the process is

\[ y_t = c + \phi y_{t-1} + \epsilon_t, \quad |\phi| < 1, \quad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2). \]  

(94)

- The parameter vector to be estimated is \( \theta = (c, \phi, \sigma^2) \).

- The joint density of a sample of size \( T \), \( y_1, y_2, \ldots, y_T \), is

\[ f(y_1, y_2, \ldots, y_{T-1}, y_T; \theta). \]  

(95)

- Function (95), viewed as a function of the parameters, for given sample values \( y_1, \ldots, y_T \), is the likelihood function, denoted as \( L(\theta) \).
Exact and conditional maximum likelihood estimation

- The maximum likelihood estimator (MLE) of $\theta$ is the value for which (95) is maximized as a function of $\theta$, given the sample values $y_1, y_2, \ldots, y_T$.

- This is the likelihood function: The joint density of the sample, viewed as a function of the parameters.
Exact and conditional maximum likelihood estimation

- For two observations, we can write the joint density as\(^{18}\)

\[
 f(y_1, y_2) = f(y_1) f(y_2|y_1)
\]  
(96)

- For three observations

\[
 f(y_1, y_2, y_3) = f(y_1) f(y_2, y_3|y_1) = f(y_1) f(y_2|y_1) f(y_3|y_1, y_2).
\]  
(97)

- For a sample of size \(T\),

\[
 f(y_1, y_2, \ldots, y_{T-1}, y_T; \theta) = f(y_1) f(y_2|y_1) f(y_3|y_1, y_2) \cdots f(y_T|y_1, y_2, \ldots, y_{T-1})
\]  
(99)

\[
 = f(y_1) \prod_{t=2}^{T} f(y_t|y_1, \ldots, y_{t-1}).
\]  
(100)

\(^{18}\)Parameters are dropped in the densities to simplify the notation.
The marginal distribution of a Gaussian ARMA process is also Gaussian.

Hence for $f(y_1)$ we have

$$y_1 \sim N \left( \frac{c}{1 - \phi}, \frac{\sigma^2}{1 - \phi^2} \right), \quad (101)$$

i.e.,

$$f(y_1) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2/(1 - \phi^2)}} \exp \left\{ -\frac{(y_1 - c/(1 - \phi))^2}{2\sigma^2/(1 - \phi^2)} \right\}. $$
Exact and conditional maximum likelihood estimation

• According to (94)

\[ y_2 | y_1 \sim N(c + \phi y_1, \sigma^2), \]  

(102)

and more generally,

\[ y_t | y_1, y_2, \ldots, y_{t-1} \overset{\text{AR(1)}}{=} y_t | y_{t-1} \sim N(c + \phi y_{t-1}, \sigma^2), \]  

(103)

• Hence, for \( t = 2, \ldots, T \),

\[
f(y_t | y_1, y_2, \ldots, y_{t-1}) = f(y_t | y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right\}.
\]
Exact and conditional maximum likelihood estimation

- The density of the sample is

\[
\begin{align*}
  f(y_1, y_2, \ldots, y_T; \theta) &= f(y_1) \prod_{t=2}^{T} f(y_t|y_{t-1}),
\end{align*}
\]

and the log–likelihood is...
Exact and conditional maximum likelihood estimation

\[ \log L(\theta) = \log f(y_1, y_2, \ldots, y_T; \theta) \]

\[ = \log f(y_1) + \sum_{t=2}^{T} \log f(y_t|y_{t-1}) \]

\[ = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log \sigma^2 + \frac{1}{2} \log(1 - \phi^2) \]

\[ - \frac{(y_1 - c/(1 - \phi))^2}{2\sigma^2/(1 - \phi^2)} \]

\[ - \sum_{t=2}^{T} \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}. \]  

\section*{(105)}

\begin{itemize}
  \item (105) is the exact log–likelihood function of the Gaussian AR(1) process, which can numerically maximized with respect to \( \theta \).
\end{itemize}
Exact and conditional maximum likelihood estimation

- The extension to Gaussian AR($p$) processes with $p > 1$ is conceptually straightforward, and efficient algorithms have been developed for calculating the density of the first $p$ observations, as well as for Gaussian MA($q$) and ARMA($p$, $q$) models.\textsuperscript{19}

Exact and conditional maximum likelihood estimation

- Alternatively, we can employ conditional maximum likelihood, which, for an AR($p$), conditions on the first $p$ observations, i.e., treats these as deterministically given.

- For these observations, no density term is added to the likelihood.

- That is, in the example above ($p = 1$), $f(y_1)$ is dropped from the likelihood.

- This somewhat simplifies the calculation of the MLE, and is asymptotically equivalent to the exact MLE (and the difference is negligible for reasonably large samples).

- For Gaussian AR($p$) processes, this turns out to be equivalent to least squares estimation.
Exact and conditional maximum likelihood estimation

• For the Gaussian AR(1) process, conditional maximum likelihood estimation, i.e., conditioning on the first observation, $y_1$, amounts to maximizing

$$\sum_{t=2}^{T} \log f(y_t|y_{t-1})$$

$$= -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log \sigma^2$$

$$- \sum_{t=2}^{T} \frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}.$$

• It is straightforward to see that (106) will be maximized with respect to parameters $c$ and $\mu$ by minimizing

$$\sum_{t=2}^{T} (y_t - c - \phi y_{t-1})^2.$$  (106)
Exact and conditional maximum likelihood estimation

• That is, conditional maximum likelihood is equivalent to ordinary least squares for Gaussian AR(1) processes.

• The conditional MLE for the error variance is

\[ \hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=2}^{T} (y_t - \hat{c} - \hat{\phi}y_{t-1})^2. \]  

(107)
Thus, in the case of a Gaussian linear model, both exact and conditional maximum likelihood estimation are feasible, though the latter is a bit more convenient.

If the error distribution is not normal, then an analytical expression for the unconditional (marginal) density $f(y_1)$ is typically not available, and exact maximum likelihood estimation is not feasible.\(^\text{20}\)

This is the case, for example, if we assume that the innovations in an AR(1) process follow a Student’s $t$ distribution with $\nu$ degrees of freedom.

\(^\text{20}\) At least not without considerable effort.
Exact and conditional maximum likelihood estimation

• Then, for $2 \leq t \leq T$, we can still write the conditional densities,

$$f(y_t | y_{t-1}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma} \left\{ 1 + \frac{(y_t - c - \phi y_{t-1})^2}{\nu\sigma^2} \right\}^{-\frac{(\nu+1)}{2}},$$

and the conditional log–likelihood

$$\log \prod_{t=2}^{T} f(y_t | y_{t-1}) = \sum_{t=2}^{T} \log f(y_t | y_{t-1})$$  \hspace{1cm} (108)

can be maximized (numerically) with respect to the parameter vector

$$\theta = (c, \phi, \sigma^2, \nu).$$  \hspace{1cm} (109)

• The same is true for GARCH models, which are typically estimated by conditional ML.\textsuperscript{21}

\textsuperscript{21}The unconditional distribution of GARCH processes is not known even if the innovations are normal.
Processes With a Unit Root

- Many economic and financial time series are nonstationary due to stochastic and/or deterministic trend components.

- Such series can be modeled with integrated ARMA (ARIMA) models.

- This refers to the situation when, in the ARMA model,

\[
Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = c + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \\
(1 - \phi_1 L - \cdots - \phi_p L^p)Y_t = c + (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t \\
\phi(L)Y_t = c + \theta(L)\epsilon_t, \quad (110)
\]

the characteristic polynomial

\[
\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad (111)
\]

has at least one root equal to 1, i.e., \(\phi(1) = 0\).
Processes With a Unit Root: Random Walk

- The simplest example is the AR(1) process with $\phi = 1$, referred to as a random walk (RW),

$$Y_t = c + Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(\sigma^2). \quad (112)$$

- Equation (112) represents a random walk with drift.\(^{22}\)

- With starting value $Y_0 = 0$,

$$Y_t = t \cdot c + (\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t), \quad (113)$$

i.e., shocks do not die out over time but accumulate (i.e., shocks are permanent).\(^{23}\) and

$$E(Y_t) = t \cdot c, \quad (114)$$

$$\text{Var}(Y_t) = \text{Var}(\epsilon_1 + \cdots + \epsilon_t) = t\sigma^2. \quad (115)$$

\(^{22}\)The case without drift corresponds to $c = 0$.

\(^{23}\)Compare with the case $|\phi| < 1$ (cf. 68), i.e., $Y_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$, where the impact of shocks on future $Y_t$s dies out at rate $\phi$.  

94
Processes With a Unit Root

- The moments depend on time, thus the random walk is not stationary.

- The next slide shows a random walk without drift \((c = 0\) in (112)) and \(\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)\).

- Note that \(Y_t\) “wanders around” in an unpredictable fashion and does not fluctuate around a constant level.
random walk $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$
Processes With a Unit Root

- If $c \neq 0$ (random walk with drift), then, from (114), the behavior of the process will be dominated by the trend component (114) over longer time horizons.\textsuperscript{24}

\textsuperscript{24}This is what one will typically expect for stock prices.
random walk with drift $y_t = 0.025 + y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$
Processes With a Unit Root

• To see that the drift component (114) dominates the increasing variance (115) over long time spans, assume $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$.

• Then, from (113), (114), and (115),

$$Y_t \sim N(ct, t\sigma^2),$$  \hspace{2cm} (116)

i.e., with probability $1 - \alpha$, $Y_t$ will be covered by the interval

$$(ct - \sigma \sqrt{t z_{1-\alpha/2}}, ct + \sigma \sqrt{t z_{1-\alpha/2}}),$$  \hspace{2cm} (117)

where $z_\alpha$ is the standard normal $\alpha$–quantile.

• The width of the interval expands more slowly than the mean function.
Processes With a Unit Root

- Nonstationarity can often be detected just from a plot of the time series, or by inspection of the sample autocorrelations.\(^{25}\)

- E.g., for the AR(1), \(Y_t = \phi Y_{t+1} + \epsilon_t\), the autocorrelation function (ACF)

\[
\text{Corr}(Y_t, Y_{t-\tau}) = \rho(\tau) = \phi^\tau.
\]

- If \(\phi\) approaches 1, with \(\phi = 1 - \delta\), then, for small \(\delta\),

\[
\rho(\tau) = \phi^\tau = (1 - \delta)^\tau = \exp\{\tau \log(1 - \delta)\} \\
\approx \exp\{-\tau\delta\} \\
\approx 1 - \tau\delta,
\]

i.e., for moderate \(\tau\), the ACF will be close to unity and decay rather slowly and in an almost linear fashion.

\(^{25}\)Formal tests for a unit root referred to as unit root tests, with the Dickey–Fuller (DF) tests being the most popular, cf. Enders (2010), Hamilton (1994, Ch. 17), and Mills and Markellos (2008, Ch. 3).
Processes With a Unit Root

- In practice, the sample ACF of a series generated by a nonstationary ARMA process will display a similar behavior.
log of S&P 500 index, 1990–2010

SACF of log of S&P 500 index
Random Walk: Transformation to Stationarity

• If $Y_t$ is a random walk,

$$Y_t = c + Y_{t-1} + \epsilon_t,$$  \hspace{1cm} (119)

then the first difference

$$\Delta Y_t = Y_t - Y_{t-1} = c + \epsilon_t$$

is a stationary (and indeed uncorrelated) process; $Y_t$ is *difference stationary*.

• Thus, if the log–price $\log P_t$ of an asset follows a random walk, then we obtain stationarity by looking at log–returns

$$r_t = \log P_t - \log P_{t-1} = \Delta \log P_t.$$
Processes with a unit root

• By means of market efficiency, it is often argued that many asset (log–) prices (such as stocks) should follow (or be very close to) a random walk.

• In this case, the log–returns $r_t = \Delta \log P_t$ should display no significant autocorrelation (be linearly unpredictable).

• However, even if asset prices follow a unit root process, this does not imply that they follow a random walk, or that returns are uncorrelated (linearly unpredictable).

• This is because the price process may be a more general unit root process, i.e., an ARIMA process.
\textbf{ARIMA}(p, d, q) Processes

• Consider the ARMA process

\[ \phi(L)Y_t = c + \theta(L)\epsilon_t. \]  \hfill (120)

• Suppose that the autoregressive polynomial \( \phi(L) \) has a unit root; then, by factoring out the unit root, we may write (120) as

\[ \phi^*(L)(1 - L)Y_t = c + \theta(L)\epsilon_t, \]  \hfill (121)

where \( \phi^*(L) = \phi(L)/(1 - L) \).

• If there is no further unit root in \( \phi^*(L) \) (and no explosive root), then (121) implies that

\[ (1 - L)Y_t = Y_t - Y_{t-1} =: \Delta Y_t \]

follows a stationary ARMA process, where \( \Delta := 1 - L \) is the difference operator.
ARIMA\( (p, d, q) \) Processes

- Thus, the first difference of an ARMA process with a simple unit root is a stationary ARMA process, where the order of the autoregressive polynomial is reduced by one.

- If there are \( d \) unit roots in \( \phi(L) \), then we can factor out \( d \) unit root and write
  \[
  \phi^*(L)(1 - L)^dY_t = c + \theta(L)\epsilon_t,
  \]
  and
  \[
  (1 - L)^dY_t = \Delta^dY_t
  \]
  is a stationary ARMA process, i.e., we have to difference \( d \) times to obtain stationarity.

- E.g., if \( d = 2 \), then
  \[
  \Delta^2Y_t = (1 - L)^2Y_t = (1 - 2L + L^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}
  \]
  will be stationary.
ARIMA($p, d, q$) Processes

- Such a process is referred to as an autoregressive integrated moving average process, denoted ARIMA($p, d, q$).

- Here $p$ refers to the number of autoregressive lags not counting the unit roots, $d$ is the number of unit roots (or the order of integration), and $q$ is the number of moving average lags.

- $Y_t$ described by an ARIMA($p, d, q$) is also referred to as being integrated of order $d$, written
  \[ Y_t \sim I(d). \]

- In practice, $d$ is typically found to be either zero (stationarity) or one, or, occasionally, two.

- The random walk is ARIMA(0,1,0).
ARIMA\((p, d, q)\) Processes

- For example, consider the AR(2) process

\[ Y_t = 1.3Y_{t-1} - 0.3Y_{t-2} + \epsilon_t. \]

- The roots of the characteristic equation satisfy

\[ \phi(z) = 1 - \phi_1z - \phi_2z^2 = 1 - 1.3z + 0.3z^2 = 0, \]

i.e.,

\[ z_{1/2} = \frac{1.3}{0.6} \pm \frac{\sqrt{1.3^2 - 0.3 \times 4}}{0.6} = \left\{ 1, \frac{1}{0.3} \right\}, \]

that is, we have one unit root an one stable root ("outside the unit circle").

- This is an ARIMA\((1,1,0)\) process.
**ARIMA**$(p, d, q)$ Processes

- We can write this process as

\[
(1 - 0.3L)(1 - L)Y_t = \epsilon_t
\]

\[
(1 - 0.3L)\Delta Y_t = \epsilon_t
\]

\[
\Delta Y_t = 0.3\Delta Y_{t-1} + \epsilon_t.
\]

- Differencing makes this process stationary, but the first differences follow an AR(1) with autoregressive parameter 0.3, i.e., they are still autocorrelated and linearly predictable.
The question as to whether asset prices follow a random walk has attracted considerable interest in the literature.

Basicallly the goal is to test whether returns are uncorrelated.

This is complicated by the fact that, as discussed above, the limiting $N(0, 1)$ distribution of the scaled sample autocorrelation coefficients, $\sqrt{T}\hat{\rho}(\tau)$, as well as the limiting $\chi^2$ distribution of the Ljung–Box statistic are valid only under the null of independent (strict) white noise.

As we have seen, this is unrealistic for most financial returns due to volatility clustering (volatility is autocorrelated).
Random walk hypothesis

• With conditional heteroskedasticity, the asymptotic standard errors of the sample ACF can be considerably larger than $1/\sqrt{T}$, leading to wider confidence intervals and larger critical values for the Box–Ljung statistic (see Taylor, 2005, for discussion).

• Thus only the null of independent white noise can be rejected based on the (convenient) standard asymptotic theory.

• For further discussion and alternative tests, see Campbell, Lo, and MacKinlay (1997, Chapter 2).\textsuperscript{26}

\textsuperscript{26} The Econometrics of Financial Markets, Princeton University Press.