Università Commerciale
Luigi Bocconi

Lecture 2: Univariate Time Series Analysis: Conditional and Unconditional Densities, Stationarity, ARMA Processes

Prof. Massimo Guidolin

20192– Financial Econometrics

Spring/Winter 2016

Overview

- Motivation: what is time series analysis
- The ARMA class
- Stationarity of ARMA models
- White noise processes
- Moving average processes
- Stationary autoregressive processes
- Wold's decomposition theorem
- Properties of AR(p) processes
- Maximum likelihood estimation of ARMA models
- Hints to Box-Jenkins' approach
- Deterministic vs. stochastic trends

Motivation: time series analysis

- Time series analysis exploits properties of past data to predict their future (density or moments)
- A time series is a sequence $\{y_1, y_2, ..., y_T\}$ or $\{y_t\}$, t=1,...,T, where t is an index denoting the period in which *x* occurs

- Can we do better than using the constant as the only predictor for financial returns?
  - Selecting time-varying predictors requires using properties of observed data to predict future observations
- Time series is branch of econometrics that deals with this question
- We shall consider univariate and multivariate time series models
  - Univariate models = relevant information set to predict one variable is restricted to the past history of that variable
- A time series is a sequence $\{y_1, y_2, ..., y_T\}$ or $\{y_t\}$, t=1,...,T, where t is an index denoting the period in time in which x occurs
  - Returns on financial assets observed over a given sample constitute the typical time series of our interest

Univariate time series processes

- Under a Gaussian IID (CER) model there is no predictability either in the mean or in the variance

- x_t is a random variable $\Rightarrow$ a time series is a sequence of random variables ordered in time; such a sequence is also known as a stochastic process
- The probability structure of a sequence of random variables is determined by the joint distribution of a stochastic process.
- The famous (Gaussian) IID/CER model is the simplest case of a probability model for such a joint distribution:

$$y_{t+1} \equiv R_{t+1} = \mu + \sigma z_{t+1} \qquad z_{t+1} \sim N(0, 1)$$

  - It implies y_{t+1} is normally IID over time with constant variance and mean equal to μ
  - CER = Constant Expected Return, R_t is the sum of constant + a white noise process, z_t
  - Under CER, forecasting is not interesting as the best forecast for the moments would be their unconditional moments

The autoregressive moving average (ARMA) class

- ARMA models are linear combinations of white noise processes:

MA(q)
$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

AR(p)
$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

ARMA(p, q)
$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t$$

$$\mathrm{E}(u_t) = 0; E\left(u_t^2\right) = \sigma^2; E(u_t u_s) = 0, t \neq s$$

- Because we have a suspicion that financial data may contain some predictability, we construct more realistic models than Gaussian IID
- In univariate strategies the basic idea is to use combinations of white noise processes, u_t, to generate more flexible models capable of replicating the relevant features of the data
- In particular, autoregressive moving average (ARMA) models are built by taking linear combinations of white noise processes

The autoregressive moving average (ARMA) class

Oh I care for this variable!

Value of Y_t absent other effects

Marginal effect of first lag

Marginal effect of first variable, increasing lags

$$Y_t = \mu + \underbrace{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}}_{p \text{ autoregressive lagsC}}$$

For instance, FTSE MIB daily returns

Let me look for variables that explain it

Omitted variables, miss-pecifications, measurement errors

$$+ \ldots + \overbrace{\theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}}^{q \text{ autoregressive lags}} + \epsilon_t$$

Most recent shocks to MIB returns

q lagged shocks to MIB returns

$\epsilon_t \; IID \; D(0, \sigma^2)$

Stationarity of ARMA Models

A series is strictly stationary if its distribution remains the same as time progresses

A weakly stationary process should have a constant mean, a variance and autocovariance structure, i.e., time-homogeneous

- Before describing the structure of and estimation approach to ARMA, one needs to define and impose on them an important statistical property that keeps ARMA series "well-behaved"
- A strictly stationary process is one where, for any $t_1, t_2, \ldots, t_T \in \mathbb{Z}$ (set of integers) and any $k \in \mathbb{Z}$ and T = 1, 2, . . ., then

  $$F y_{t_1}, y_{t_2}, \ldots, y_{t_T}(y_1, \ldots, y_T) = F y_{t_1+k}, y_{t_2+k}, \ldots, y_{t_T+k}(y_1, \ldots, y_T)$$

  where $F(\bullet)$ denotes the joint CDF
  - The probability measure for the sequence $\{y_t\}$ is the same as that for {yt+k} $\forall k$ ('$\forall k$' means 'for all values of k')
- A weakly stationary process satisfies instead 3 properties $\forall t \geq 1$:

$$E(y_t) = \mu \qquad E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty \qquad E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1} \quad \forall\, t_1, t_2$$

Stationarity of ARMA Models

The moment $E[(y_t - E(y_t)(y_{t-s} - E(y_{t-s}))] = \gamma s$, s = 0, 1, 2, ... is the **autocovariance function**

When $s = 0$, the autocovariance at lag zero is the variance of y

- All these moments do not change over the history of a time series
- The autocovariances determine how y is related to its previous values, and for a stationary series they depend only on the difference between t_1 and t_2, so that the covariance between y_t and y_{t-1} is the same as the covariance between y_{t-10} and y_{t-11}, etc.
- The autocovariances are not a particularly useful measure because their values depend on the units of measurement of y_t

- It is more convenient to use the **autocorrelations**, which are the autocovariances normalised by dividing by variance $\tau_s = \gamma_s/\gamma_0$
  - If τ_s is plotted against s = 0, 1, 2, . . . , the **autocorrelation function** (acf) or autocorrelogram is obtained
- The ACF allows us to provide an intuitive definition of **white noise process, one with no discernible structure**

From White Noise to Moving Average Processes

A white noise process is one with no serial correlation structure

A moving average model is simply a linear combination of white noise processes

- The u_t shock of models is white noise
- A white noise process is such that:
- Has zero autocovariances, except at lag 0

$$E(y_t) = \mu$$
$$\text{var}(y_t) = \sigma^2$$
$$\gamma_{t-r} = \begin{cases} \sigma^2 & \text{if} \quad t = r \\ 0 & \textit{otherwise} \end{cases}$$

- The simplest class of time series model that one could entertain is that of the moving average process:

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$$

  where u_t is white noise
- When q = 0, then a MA(0) reduces to a white noise process
  - In much of what follows, the constant (μ) is dropped from the equations because it considerably eases the algebra involved
- The distinguishing properties of the MA(q) are as follows

Moving Average Processes

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

(2) $\mathrm{var}(y_t) = \gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2$

(3) covariances γ_s

$$= \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s})\,\sigma^2 & \mathit{for}\;\; s = 1, 2, \ldots, q \\ 0 & \mathit{for}\;\; s > q \end{cases}$$

- A MA(q) process has constant mean and variance, and autocovariances which may be non-zero to lag q and will be 0 thereafter
  - For instance, consider the MA(2): $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
  - Let's see why the properties above hold in this case
  - If $\mathrm{E}(u_t) = 0$, then $\mathrm{E}(u_{t-i}) = 0\ \forall i$ and so

$$\mathrm{E}(y_t) = \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})$$
$$= \mathrm{E}(u_t) + \theta_1 \mathrm{E}(u_{t-1}) + \theta_2 \mathrm{E}(u_{t-2}) = 0$$

  - Moreover

$$\mathrm{var}(y_t) = \mathrm{E}[y_t - \mathrm{E}(y_t)][y_t - \mathrm{E}(y_t)] = \mathrm{E}[(y_t)(y_t)]$$
$$= \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})]$$

Moving Average Processes

(1) $E(y_t) = \mu$

(2) $\text{var}(y_t) = \gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2$

(3) covariances γ_s

$$= \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s})\,\sigma^2 & \textit{for} \quad s = 1, 2, \ldots, q \\ 0 & \textit{for} \quad s > q \end{cases}$$

$$\text{var}(y_t) = E\left[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 + \textit{cross-products}\right]$$

- But $E[\textit{cross-products}] = 0$ because $\text{cov}(u_t, u_{t-s}) = 0$ for $s \neq 0$
- One does not need to worry about these cross-product terms, since these are effectively the autocovariances of u_t, which will all be zero
- Therefore

$$\text{var}(y_t) = \gamma_0 = E\left[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2\right] = \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 = \left(1 + \theta_1^2 + \theta_2^2\right)\sigma^2$$

- The autocovariance at lag 1 is computed as:

$$\gamma_1 = E[y_t][y_{t-1}]$$

$$\gamma_1 = E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3})]$$

Moving Average Processes

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

(2) $\mathrm{var}(y_t) = \gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2$

(3) covariances γ_s

$$= \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s})\,\sigma^2 & \textit{for} \quad s = 1, 2, \ldots, q \\ 0 & \textit{for} \quad s > q \end{cases}$$

- Again, ignoring cross-products, this can be written as:

$$\gamma_1 = \mathrm{E}\left[\left(\theta_1 u_{t-1}^2 + \theta_1\theta_2 u_{t-2}^2\right)\right]$$
$$\gamma_1 = \theta_1\sigma^2 + \theta_1\theta_2\sigma^2$$
$$\gamma_1 = (\theta_1 + \theta_1\theta_2)\sigma^2$$

- The autocovariances at lags 2-3 are:

$$\gamma_2 = \mathrm{E}[y_t][y_{t-2}]$$
$$\gamma_2 = \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4})]$$
$$\gamma_2 = \mathrm{E}\left[\left(\theta_2 u_{t-2}^2\right)\right]$$
$$\gamma_2 = \theta_2\sigma^2$$

$$\gamma_3 = \mathrm{E}[y_t][y_{t-3}]$$
$$\gamma_3 = \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-3} + \theta_1 u_{t-4} + \theta_2 u_{t-5})]$$
$$\gamma_3 = 0$$

- So $\gamma_s = 0$ for s > 2, as claimed

Autoregressive Processes

An autoregressive model is one where the current value of y depends upon only values in previous period plus an error term

- Finally, because the variance is just γ_0, the autocorrelations will be:

$$\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1 + \theta_1\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{(\theta_1 + \theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)} \quad \tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_2}{(1+\theta_1^2+\theta_2^2)} \quad \tau_3 = \frac{\gamma_3}{\gamma_0} = 0$$

- A sample ACF is plotted

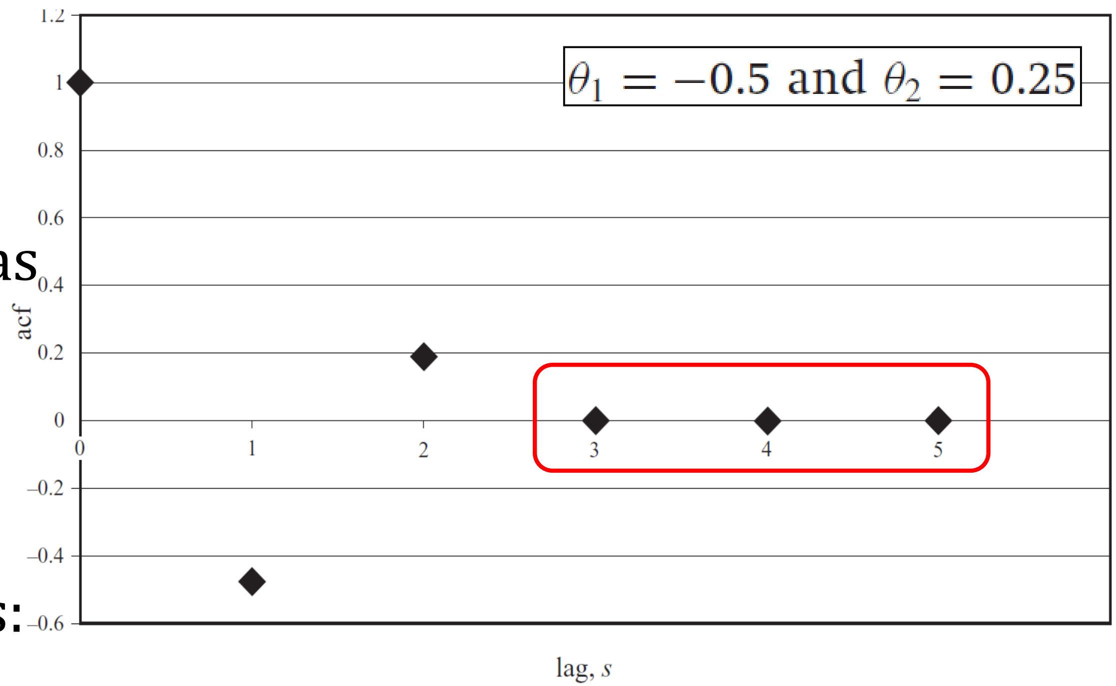

Figure 5.1 Autocorrelation function for sample MA(2) process

- An autoregressive model of order p, denoted as AR(p), can be expressed as

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

where u_t is white noise

- Equivalently, the model is:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

Stationary Autoregressive Processes

The stationarity of an autoregressive model depends on the (estimated) values of its parameters in a MA-type representation

- Stationarity is a desirable property of an estimated AR model
- A model whose coefficients are nonstationary has the unfortunate property that previous values of the error term will have a non-declining effect on the current value of y_t as time progresses
- To think of non-stationarity it is therefore useful to try and re-write the model in terms of a moving average representation
  - Let's consider an AR(2) process and set $\mu = 0$ to keep algebra simple
  - $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t \;\Rightarrow\; y_t = \phi_1(\phi_1 y_{t-2} + \phi_2 y_{t-3} + u_{t-1}) + \phi_2 y_{t-2} + u_t$

$$= [(\phi_1)^2 + \phi_2] y_{t-2} + \phi_1\phi_2 y_{t-3} + \phi_1 u_{t-1} + u_t$$

$$= [(\phi_1)^2 + \phi_2](\phi_1 y_{t-3} + \phi_2 y_{t-4} + u_{t-2}) + \phi_1\phi_2 y_{t-3} + \phi_1 u_{t-1} + u_t$$

$$\Rightarrow\; y_t = \{[(\phi_1)^2 + \phi_2] + \phi_2\}\phi_1 y_{t-3} + [(\phi_1)^2 + \phi_2]\phi_2 y_{t-4} + [(\phi_1)^2 + \phi_2]u_{t-2} + \phi_1 u_{t-1} + u_t$$

    which shows that by recursive backward substititution an AR(2) process is becoming a MA(q) one
  - But one can go on forever: the AR(p) becomes a MA(∞)

Stationary Autoregressive Processes

A stationary AR(p) process can be written as an MA(∞) and this implies that restrictions need to be imposed on the coefficients

- Such a condition however holds if and only if the weights attached to lagged values for *y* becomes less and less important as the backward recursive substitution process goes on
  - This depends on the factors that in the backward substitution process have been developing as $\{[(\phi_1)^2 + \phi_2]+\phi_2\}\phi_1$, $[(\phi_1)^2+ \phi_2]\,\phi_2$, etc.
- Adequate conditions will be needed for the AR(*p*) coefficients to force the weights attached to y_{t-Q} and y_{t-Q+1} to disappear as $Q \rightarrow \infty$
- Intuitively, this means that the autocorrelations involving y_t, y_{t-Q} and y_{t-Q+1} will decline eventually as the lag length is increased
- Because such coefficients attached to long-dated, remote lags of *y*, will eventually disappear, **a stationary AR(*p*) process can be written as an MA(∞)**
- **If the process is stationary, the coefficients in the MA(∞) representation will decline eventually with lag length**

Stationary Autoregressive Processes

A general AR(p) model is weakly stationary iff the roots of the "characteristic equation" all lie strictly outside the unit circle

- The condition for testing the weak stationarity of a general AR(p) model is that the roots of the "characteristic equation"

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

  all lie strictly outside the unit circle
  - The characteristic equation determines the behavior of the ACF as a function of the lag number
  - For instance, consider $y_t = y_{t-1} + u_t$; the characteristic equation is 1-1xz = 0, which has a unique root of z* = 1, that fails to lie outside the unit circle; hence $y_t = y_{t-1} + u_t$ turns out to be a nonstationary process
    - Under the efficient market hypothesis (EMH), we expect asset prices to follow a non predictable process and hence to be non-stationary
  - $y_t = y_{t-1} + u_t$ is known as a driftless random walk process
  - See example 6.3 in Brooks to have a less trivial example that however implies using the lag operator, some L such that $L^Q y_t = y_{t-q}$

Wold's Decomposition Theorem

Wold's theorem states that any weakly stationary series can be decomposed into the sum of two unrelated processes, a purely deterministic part and a stochastic, MA(∞) part

- The characterization just examined represents an application of Wold's decomposition theorem
- The theorem states that any stationary series can be decomposed into the sum of two unrelated processes, a purely deterministic part and a purely stochastic part, which will be an MA(∞)
- This result is crucial to derive the ACF of the process
- The characteristics of an autoregressive process are as follows:

$$E(y_t) = \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

- For instance, in the case of an AR(1) process,

$$\mathrm{E}(y_t) = \mathrm{E}(\mu + \phi_1 y_{t-1}) \quad \underset{y_{t-1} = \mu + \phi_1 y_{t-2} + u_{t-1}}{\overset{y_t = \mu + \phi_1 y_{t-1} + u_t}{\Rightarrow}} \quad \mathrm{E}(y_t) = \mu + \phi_1(\mu + \phi_1 \mathrm{E}(y_{t-2}))$$

$$\mathrm{E}(y_t) = \mu + \phi_1 \mathrm{E}(y_{t-1}) \qquad \mathrm{E}(y_t) = \mu + \phi_1 \mu + \phi_1^2 \mathrm{E}(y_{t-2})$$

Properties of Autoregressive Processes

$$\mathrm{E}(y_t) = \mu + \phi_1(\mu + \phi_1 \mathrm{E}(y_{t-2})) \quad y_{t-1} = \mu + \phi_1 y_{t-2} + u_{t-1} \quad \mathrm{E}(y_t) = \mu + \phi_1\mu + \phi_1^2(\mu + \phi_1 \mathrm{E}(y_{t-3}))$$

$$\Rightarrow$$

$$\mathrm{E}(y_t) = \mu + \phi_1\mu + \phi_1^2 \mathrm{E}(y_{t-2}) \quad y_{t-2} = \mu + \phi_1 y_{t-3} + u_{t-2} \quad \mathrm{E}(y_t) = \mu + \phi_1\mu + \phi_1^2\mu + \phi_1^3 \mathrm{E}(y_{t-3})$$

- Now a pattern emerges and making such n substitutions, we obtain

$$\mathrm{E}(y_t) = \mu\left(1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{n-1}\right) + \phi_1^n \mathrm{E}(y_{t-n})$$

- So long as the model is stationary, i.e., $|\varphi_1| < 1$, then $\lim_{n\to\infty}|\varphi_1|^n = 0$
- Therefore, $\lim_{n\to\infty}\varphi_1^n \mathrm{E}(y_{t-n}) = 0$, and so $\mathrm{E}(y_t) = \mu\left(1 + \phi_1 + \phi_1^2 + \cdots\right)$
- Because $|\varphi_1| < 1$ guarantees that the geometric series converges to the sum $(1 - \varphi_1)$, $$E(y_t) = \frac{\mu}{1 - \phi_1}$$

- Autocovariances and autocorrelation functions are obtained by solving a set of simultaneous equations, the Yule-Walker equations
  - The Yule--Walker equations express the correlogram (the τs) as a function of the auto-regressive coefficients (the φs)
  - Therefore when p = 1 so that $\phi_2 = \phi_3 = \ldots = \phi_p = 0$ then

$$\begin{aligned}
\tau_1 &= \phi_1 + \tau_1\phi_2 + \cdots + \tau_{p-1}\phi_p \\
\tau_2 &= \tau_1\phi_1 + \phi_2 + \cdots + \tau_{p-2}\phi_p \\
&\vdots \quad \vdots \quad \vdots \\
\tau_p &= \tau_{p-1}\phi_1 + \tau_{p-2}\phi_2 + \cdots + \phi_p
\end{aligned}$$

Properties of Autoregressive Processes

- $\tau_1 = \phi_1$; $\tau_2 = \tau_1\phi_1 = (\phi_1)^2$; ..., $\tau_p = \tau_{p\text{-}1}\phi_1 = (\phi_1)^p$
- When p = 2, we have instead
  $\tau_1 = \phi_1 + \tau_1\phi_2 \Rightarrow \tau_1 = \phi_1/(1\text{-}\phi_2)$
  $\tau_2 = \tau_1\phi_1+\phi_2=(\phi_1+\tau_1\phi_2)\phi_1+\phi_2=(\phi_1)^2+\tau_1\phi_2\phi_1+ \phi_2= (\phi_1)^2+(\phi_1)^2\phi_2/(1\text{-}\phi_2)+\phi_2$
  ...
- As for the variance, Appendix A shows that: $\operatorname{var}(y_t) = \dfrac{\sigma^2}{(1-\phi_1^2)}$
- More generally var(y_t) come from the Yule-Walker equations
- Example 5.4 gives evidence of raw calculations of autocorrelations calculations for the AR(1) process, although these need to be identical to th $\gamma_1 = \mathrm{E}[y_t y_{t-1}]$ l by the Yule-Walker equations
- $$\gamma_1 = \mathrm{E}\left[\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\right)\left(u_{t-1} + \phi_1 u_{t-2} + \phi_1^2 u_{t-3} + \cdots\right)\right]$$
  $$\gamma_1 = \mathrm{E}\left[\phi_1 u_{t-1}^2 + \phi_1^3 u_{t-2}^2 + \cdots + cross - products\right]$$
  $$\gamma_1 = \phi_1\sigma^2 + \phi_1^3\sigma^2 + \phi_1^5\sigma^2 + \cdots$$
  $$\gamma_1 = \phi_1\sigma^2\left(1 + \phi_1^2 + \phi_1^4 + \cdots\right) \quad = \frac{\phi_1\sigma^2}{(1-\phi_1^2)}$$

Properties of Autoregressive Processes

$$y_t = 0.9y_{t-1} + u_t$$

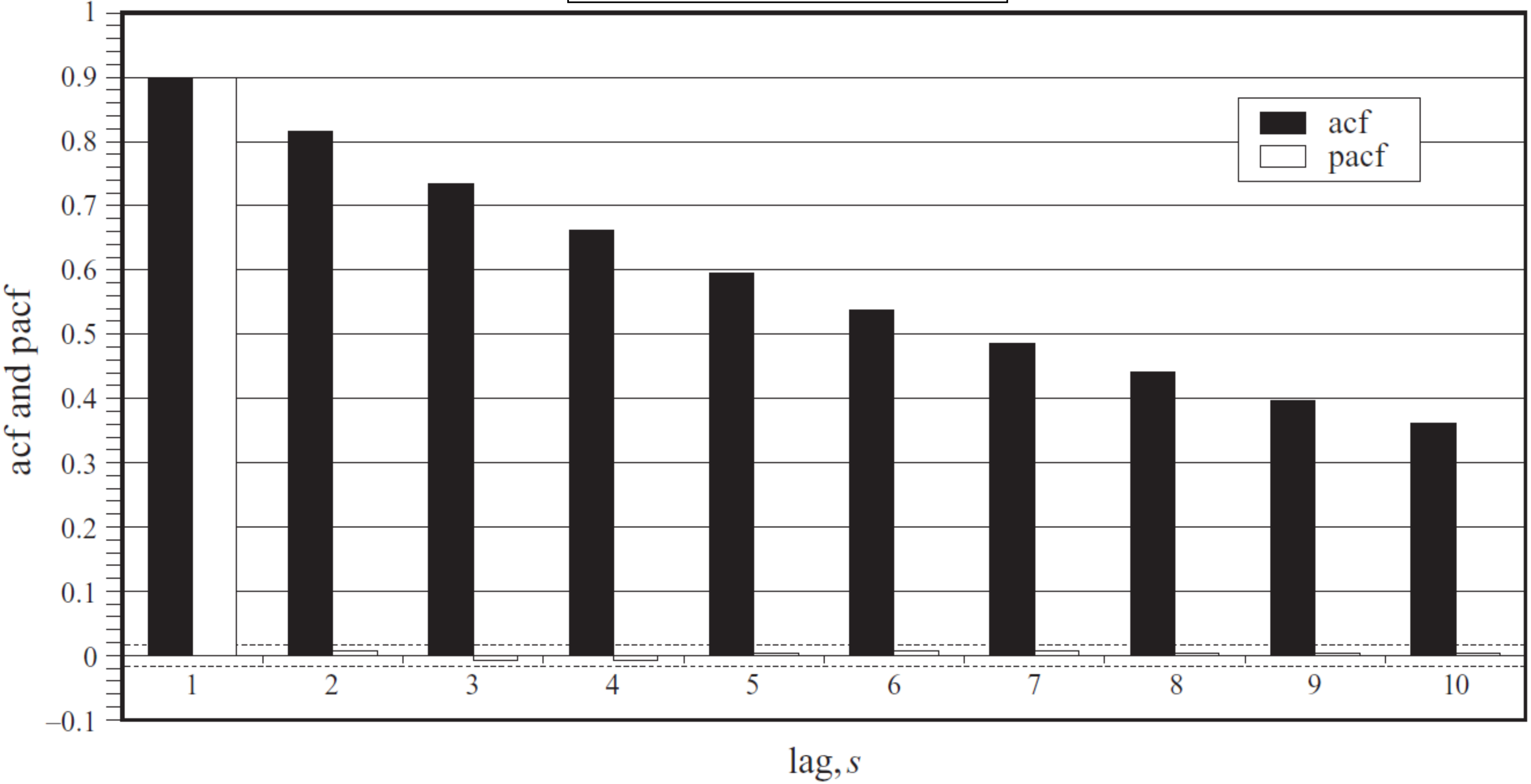

Figure 5.4 Sample autocorrelation and partial autocorrelation functions for a slowly decaying AR(1) model: $y_t = 0.9y_{t-1} + u_t$

Properties of Autoregressive Processes

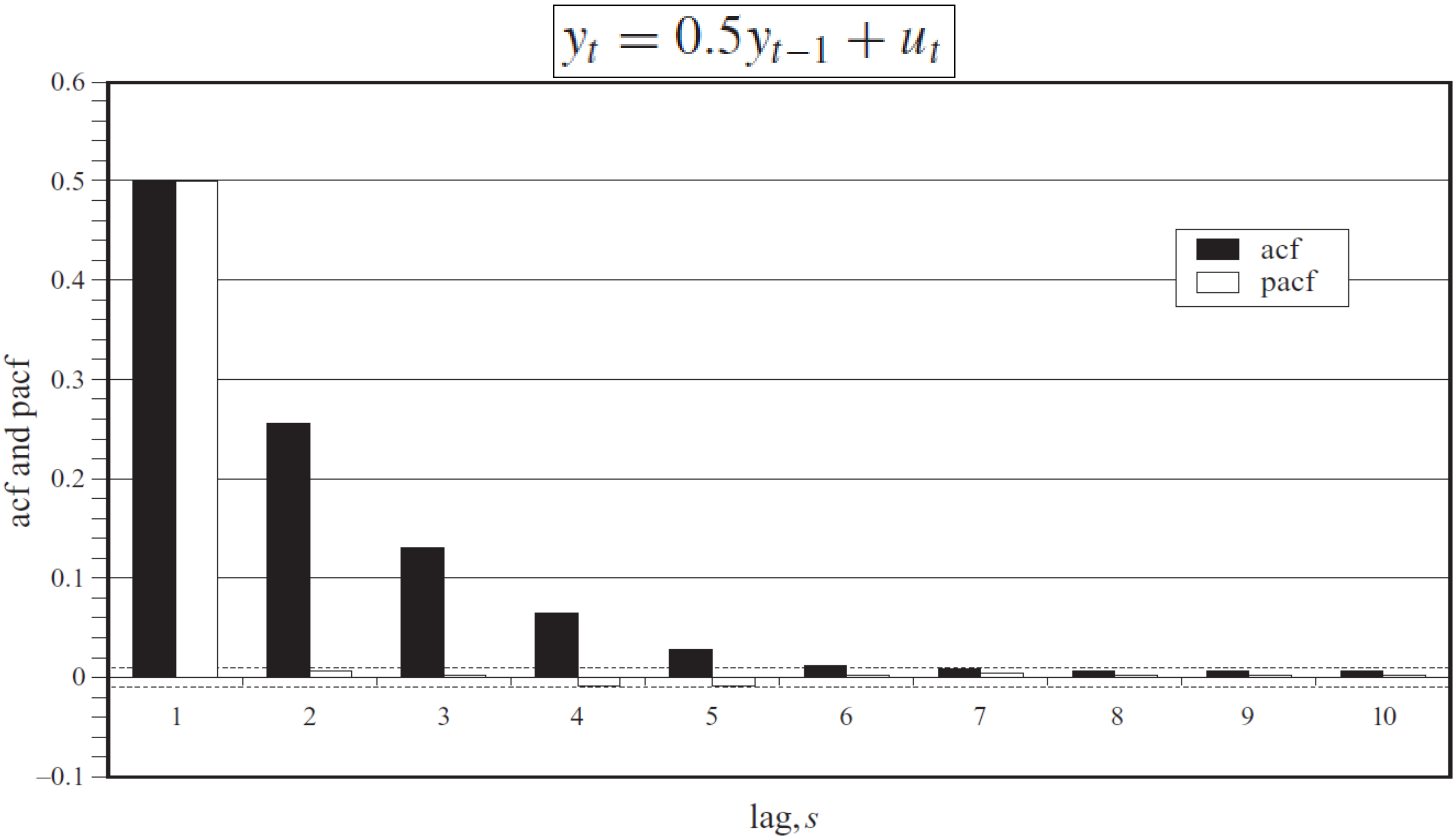

Figure 5.5 Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model: $y_t = 0.5y_{t-1} + u_t$

Properties of Autoregressive Processes

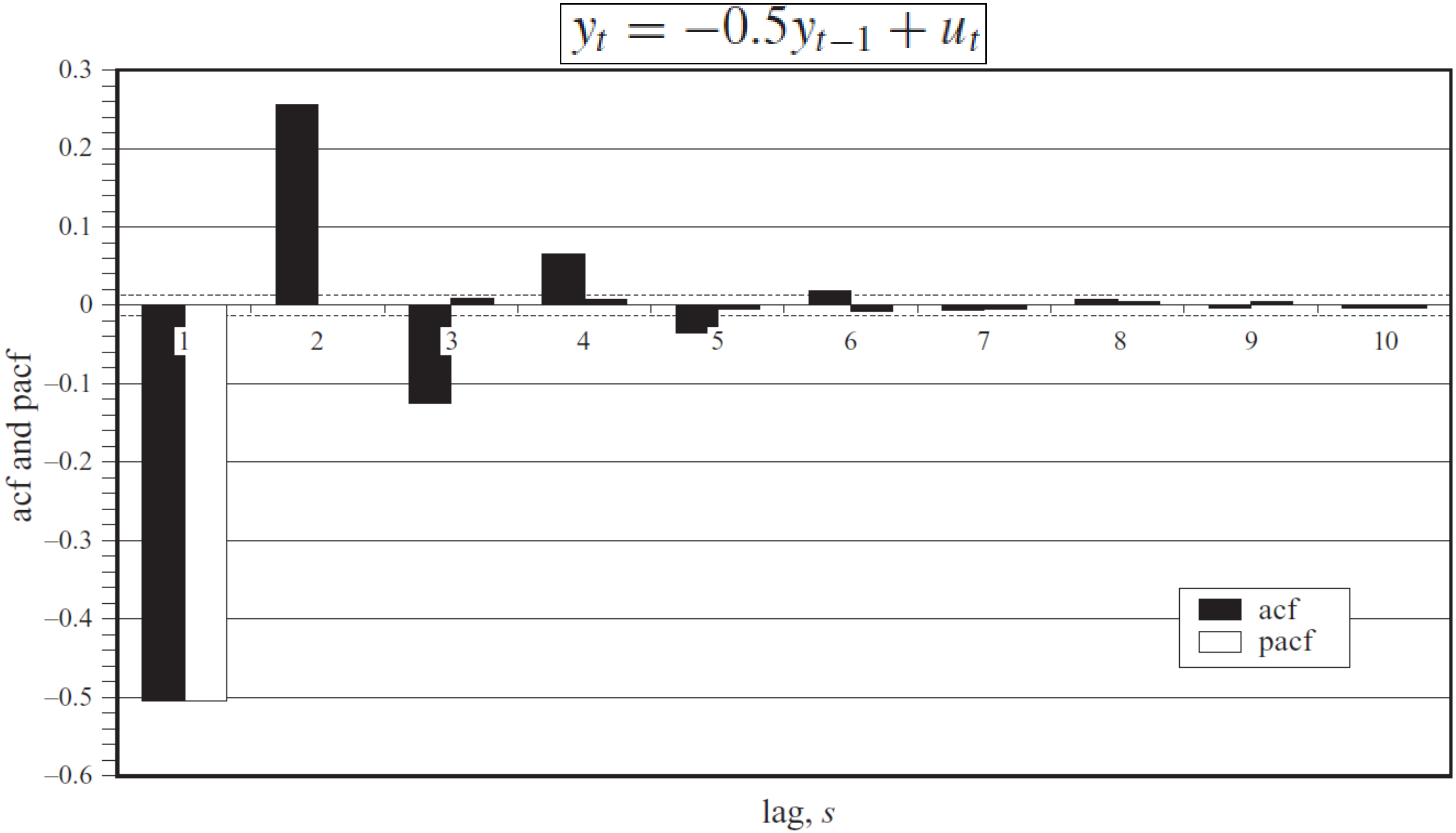

Figure 5.6 Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model with negative coefficient: $y_t = -0.5y_{t-1} + u_t$

Properties of Autoregressive Processes

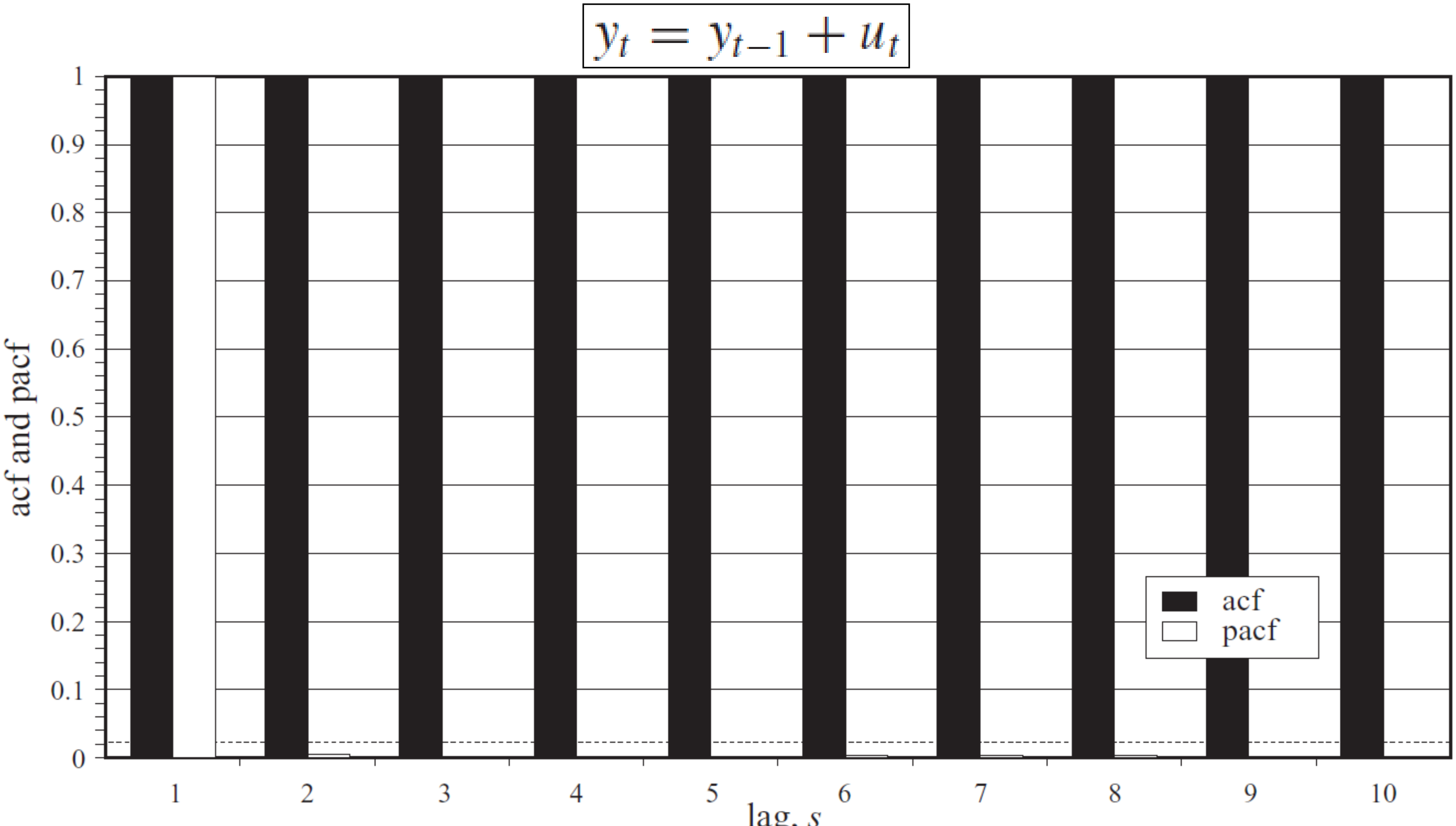

Figure 5.7 Sample autocorrelation and partial autocorrelation functions for a non-stationary model (i.e. a unit coefficient): $y_t = y_{t-1} + u_t$

Autoregressive Moving Average Processes

The characteristics of an ARMA process will be a combination of those from the AR and MA parts

- By combining the AR(p) and MA(q) models, an ARMA(p, q) model is obtained:
$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t$$
- The characteristics of an ARMA process will be a combination of those from the AR and MA parts
  - E.g., the ACF will display combinations of behaviour derived from the AR and MA parts, but for lags > q, the it will simply be identical to the individual AR(p) model, so that the AR part will dominate
- How do you estimate an ARMA model?
  - In the AR case, there is no problem: just apply standard regression analysis to linear specifications where the regressors are lags of the dependent variable
  - However, standard regression methods are no longer applicable when MA terms appear

ML estimation of ARMA models

- ML estimation is based on the maximization of the likelihood function, the joint density of all available data in the sample

- This derives from the fact that some key conditions of classical regression analysis are violated
  - In a nutshell, the standard classical linear regression conditions required by OLS in $y_t = \theta_1 u_{t-1} + u_t$ fails as u_{t-1} is a stochastic regressor
- A more general method, capable of dealing with these issues is Maximum Likelihood Estimation (MLE)
- Estimates of the parameters of interests are obtained by maximizing the likelihood function
  - The likelihood function is the joint probability distribution of the data, that depends on the observations on the time series of interest and on the unknown parameters
  - It is defined on the parameter space Θ, given the observation of the observed sample $\{y_t\}$ t=1, ..., T and of a set of initial conditions y_0
    - One can interpret such initial conditions as the pre-sample observations on the relevant variables

ML estimation of ARMA models

- Although it represents the joint density of the data, in MLE the log-likelihood is maximized choosing parameter estimates for fixed data

- Once a sample of observations is fed to the likelihood, the latter can be treated as a function of the unknown coefficients
- The MLE is then obtained by choosing the value of the unknown parameters that maximize the likelihood function
- In practice, the **MLE selects the value of parameters to maximize the probability of drawing data that are effectively observed**
- We now provide the example of the MLE of an MA(1) process:

$$r_{t+1} = \theta_0 + \varepsilon_{t+1} + \theta_1 \varepsilon_t \qquad \varepsilon_t \sim NID\left(0, \sigma^2\right)$$

  - In this case the unknown parameters to be estimated are θ_0, θ_1, and σ^2
- To derive MLEs, first define the time series of residuals:

$$\begin{aligned} \varepsilon_{t+1} &= r_{t+1} - \theta_0 - \theta_1 \varepsilon_t \\ \varepsilon_0 &= 0. \end{aligned}$$

ML estimation of ARMA models

- Under IID shocks (returns) the likelihood is obtained as the product of the density function of each random observation
- The likelihood may often be maximized only numerically

- Given the distributional assumption on ε_{t+1}, we have:

$$f\left(\varepsilon_{t+1}\right)=\frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{1/2}}\exp\left(-\frac{\varepsilon_{t+1}^{2}}{2\sigma_{\varepsilon}^{2}}\right)$$

  - This expression is the distribution of a single observation, while the likelihood function is the joint distribution of the entire sample
- If the ε_{t+1} are independent over time, then the likelihood function can be written as follows:

$$f\left(\varepsilon_{1},\varepsilon_{2},\ldots\varepsilon_{t+1}\right)=\prod_{i=1}^{T}f\left(\varepsilon_{i}\right)=\prod_{i=1}^{T}\frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{1/2}}\exp\left(-\frac{\varepsilon_{i}^{2}}{2\sigma_{\varepsilon}^{2}}\right)$$

- The MLE chooses θ_0, θ_1,σ^2_ε to maximize the probability that the estimated model has generated the observed data
  - The optimum is not always found analytically, iterative search is the standard method, easily implemented in EViews or even Excel

Hints to the Box-Jenkins approach

- The Box-Jenkins approach is a structured sequence of steps aiming at best specifying and estimating models in the ARIMA class

- What is the I in ARIMA? It stands for integrated, to mean that y_t contains a stochastic trend, or $y_{t+1} = y_t + u_{t+1}$ with u_{t+1} white noise
  - White noise = u_{t+1} is IID, no serial correlation, constant variance
- The approach is structured on FIVE STEPS:

❶ PRE-WHITENING: make sure that the time series is stationary

- Make sure that the model at hand is ARMA and not ARIMA: this is commonly achieved for simple univariate time series via differencing, i.e., by considering $\Delta y_t = (1 - L)y_t$ instead of y_t

❷ MODEL SELECTION: look for the best ARMA specification

- Information criteria are a useful tool to this end
- They are model selection criteria based on penalized versions of the maximized log-likelihood function

Hints to the Box-Jenkins approach

- The key steps of Box-Jenkins' approach are: pre-whitening; model selection (pick *p* and *q*); estimation (often by MLE); model checking/diagnostic; forecasting/use in decisions

- They are used to select *p* and *q* in an ARMA(*p*, *q*) model, e.g., Akaike's (AIC) and the Schwarz Bayesian (SBIC) are the most commonly used criteria

$$AIC \equiv -2\ln(\mathcal{L}) + 2(p+q)$$

$$BIC \equiv -2\ln(\mathcal{L}) + \ln(n)(p+q)$$

Likelihood function

Number of observations (*n*=T)

❸ ESTIMATION: see above

❹ MODEL CHECKING/DIAGNOSTICS: make sure residuals reflect any assumptions that were made, e.g., u_{t+1} IID $N(0, \sigma^2)$

❺ FORECASTING: the selected and estimated model is typically simulated forward to produce forecasts for the variable of interests at one or more relevant horizons

- Time series of long-horizon returns (computed as sums of higher frequency returns) besides being persistent, often feature trends

Deterministic vs. stochastic trends

- There are two types of trends, stochastic and deterministic
- In a stochastic trend, the baseline is a random walk, $y_{t+1} = \rho_0 + \rho_1 y_t + u_{t+1}$ to be decomposed in deterministic comp. + trend
- Stochastic integrated series are made stationary by differentiating them

---

- Two basic kinds of trends
- Stochastic ones, characterizing random walk processes (below, with drift):
  $$y_t = a_0 + y_{t-1} + u_t \qquad u_t \sim IID\ N(0, \sigma^2)$$
  - Recursive substitution yields $y_t = a_0 + a_0 t + \sum_{i=0}^{t-1} u_{t-i}$
- This shows the structure: deterministic component (a_0t) + stochastic trend, here $\sum_{i=0}^{t-1} u_{t-i}$
  - The series is non-stationary in that the unconditional mean ($E(y_t) = y_0 + a_0 t$) is a function of time

Deterministic vs. stochastic trends

- A time series that needs to be differentiated *d* times before becoming stationary, is said to be **integrated** of order *d*
- In an integrated series, all past shocks matter equally and have **permanent** effects in the infinite distant future

---

- An immediate way to make a non-stationary series stationary is by differencing it:

$$\Delta y_t \equiv y_t - y_{t-1} = (1 - L)y_t = a_0 + u_t$$

- If $\{x_t\}$ needs differentiation *d* times, it is **integrated of order d, I(d)**
  - A random walk with drift is clearly I(1)
- Assuming $a_0 = 0$, a random walk may be re-written as:

$$y_t = y_{t-1} + u_t = y_{t-2} + u_{t-1} + u_t = y_{t-3} + u_{t-2} + u_{t-1} + u_t = \ldots$$

  this means that all past shocks matter equally and have permanent effects in the infinite distant future
- In this sense, I(1) processes display maximum persistence...

Deterministic vs. stochastic trends

- A deterministic trend is a process the value of which directly depends on time (t) as a variable
- This type of non-stationarity is simply removed by regressing z_t on the deterministic trend

---

- The alternative is represented by deterministic trends:

$$z_t = \alpha + \beta t + \epsilon_t$$

- These processes are also called trend-stationary
- The process for z_t is non-stationary, but non-stationarity is removed simply by regressing z_t on the deterministic trend
- Unlike the stochastic case, for integrated processes the removal of deterministic trend does not deliver a stationary time-series

Reading List/How to prepare the exam

- Carefully read these Lecture Slides + class notes
- Possibly read BROOKS, chapter 6.
- You may want to take a look at CHRISTOFFERSEN, chapter 3.
- Lecture Notes are available on Prof. Guidolin's personal web page

Appendix A: The Variance of an AR(1) Process

- From the definition of the variance of any random variable *y*, it is possible to write

$$\text{var}(y_t) = \text{E}[y_t - \text{E}(y_t)][y_t - \text{E}(y_t)]$$

  but $\text{E}(y_t) = 0$, since μ is set to zero, so that

$$\text{var}(y_t) = \text{E}[(y_t)(y_t)]$$

$$\text{var}(y_t) = \text{E}\left[\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\right)\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\right)\right]$$

$$\text{var}(y_t) = \text{E}\left[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \cdots + \textit{cross-products}\right]$$

$$\text{var}(y_t) = \gamma_0 = \text{E}\left[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \cdots\right]$$

$$\text{var}(y_t) = \sigma^2 + \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \cdots$$

$$\text{var}(y_t) = \sigma^2 \left(1 + \phi_1^2 + \phi_1^4 + \cdots\right)$$

- Provided that $|\varphi_1| < 1$, the infinite sum can be written as

$$\text{var}(y_t) = \frac{\sigma^2}{\left(1 - \phi_1^2\right)}$$