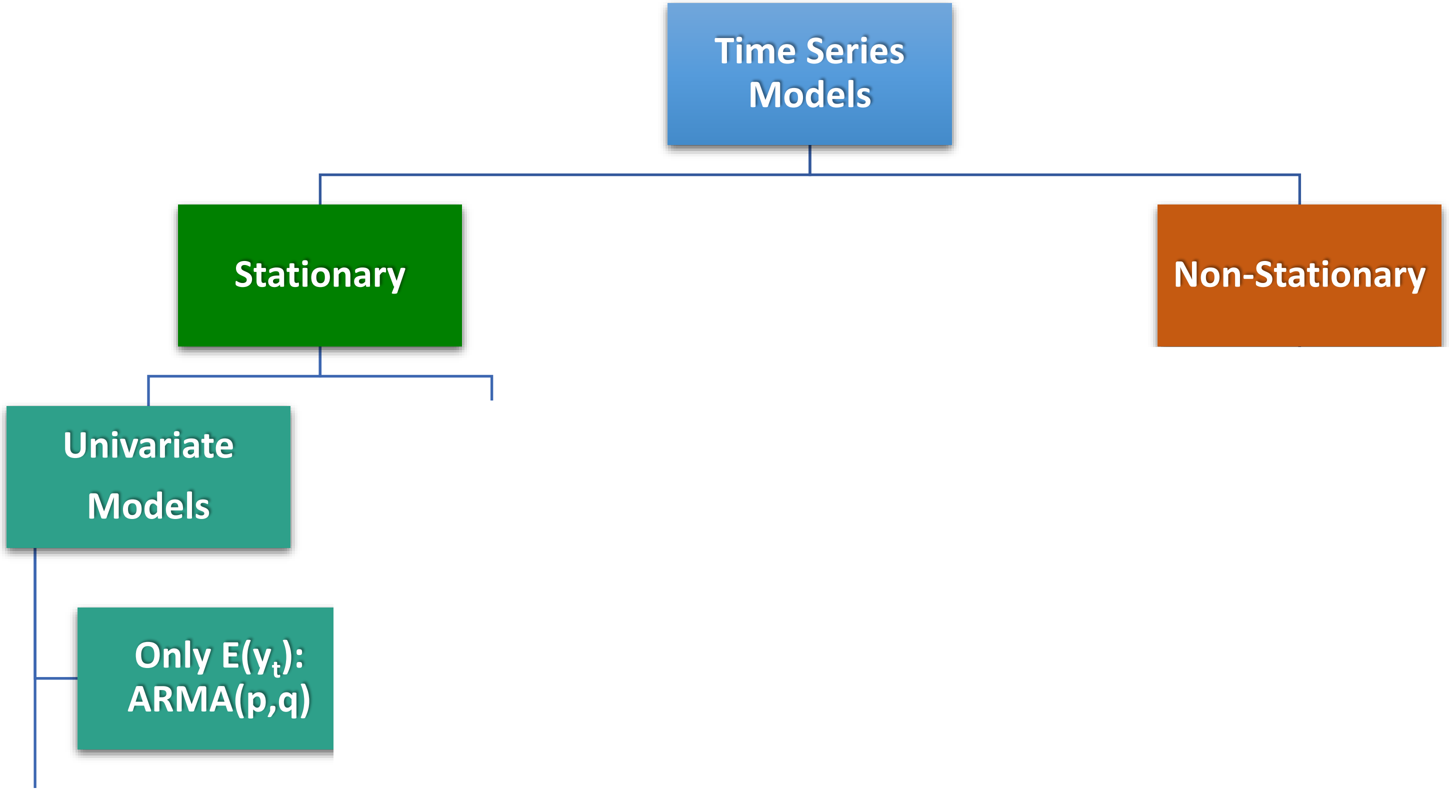
Advanced Time Series Econometrics

Topic 2: Univariate Time Series Models

Box-Jenkins analysis



Time Series Models
Stationary
Non-Stationary
Univariate Models
Only $E(y_t)$: ARMA(p,q)

Objective in Topic 2:

- Define stationarity, find moments of stationary processes
- Find the best univariate, linear, stochastic model for a given time-series
    - Univariate
        - the only observable information used is the series itself
    - Stochastic
        - Our model will not fit perfectly, have random components
    - Find
        - Clarify steps to take in a regression analysis of a single variable
    - Best
        - To be carefully defined
    - Model
        - Data generating process is not known with certainty = must be estimated

Purpose of dynamic econometrics

Almost all questions in dynamic econometrics can be reduced to the following:

- If there is a change to one variable at a point in time, what are the effects on that and other variables over time?

Examples:

  - Monetary policy: interest rates, inflation rates, output gaps
  - Fiscal policy: budget deficit, consumption, investment
  - International Finance: interest rate differentials, exchange rate, net foreign asset positions

Today we focus on the univariate version – everything extends naturally

Motivation

- We will “explain” the time path of a variable with only its own history
  - Economic theory unlikely to suggest this
  - Some theoretical models admit a univariate *reduced form*
  - This section is mostly for
    (a) historical completeness
    (b) theoretical base for general multivariate models
    (c) simplest forecasting models in a comparative study
- Directly applicable uses for a modern econometrician:
  - “Advanced Descriptive Statistics”
    - Splits time series information into
      - Systematic and Stochastic
      - Explained and Unexplained
  - Allows naïve forecasting and probabilistic claims - Still used as one of the benchmarks of a more complicated forecasting model

Definitions: your turn

- Provide a definition for the following concept:
  a real valued random variable

Definitions

- Real valued random variable
  - A function that maps the set of possible events to the real line
  - Loosely: A mathematical object that can take on more than one real value but whose outcome cannot be perfectly predicted
  - Random variable x is fully described by its probability distribution function
$$F(z) \equiv \Pr(x \leq z)$$
  - We summarize the features of a probability distribution function by considering its moments

Definitions (2)

- Moments of a distribution?

Definitions (2)

- First four (centered) moments of a distribution:
  - The Expected value
  - Variance
  - Skewness
  - Kurtosis
- First two moments of a time series:
  - Expected Value
  - Variance and auto-covariances

From difference equations to univariate time series models

- Difference equations and the theories surrounding them and their solutions are the mathematical foundations that we use in time series
- From now on, we will only consider these difference equations as statistical models of the underlying data generating process we are trying to describe

Univariate time series models

- Univariate linear time series are modelled as stochastic difference equations
- Typically an Auto-regressive (AR), moving average (MA) process: E.g. an ARMA(2,3) process

$$y_t = a_0 + \underbrace{a_1 y_{t-1} + a_2 y_{t-2}}_{\text{Auto-Regressive of order 2}} + \overset{\text{White Noise}}{\varepsilon_t} + \underbrace{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}}_{\text{Moving Average of order 3}}$$

Univariate time series models

- Univariate linear time series are modelled as stochastic difference equations
- To be precise: we postulate that the true data generating process follows an ARMA process, and then attempt to find the best estimate
- The central concept that distinguishes different approaches/methods applicable to time series models is **stationarity.**

Stationarity

A time series process y_t is called *weakly stationary* (or *covariance stationary*) if:

$\forall\, t, s, j$:

$$\mathbb{E}(y_t) = \mathbb{E}(y_{t-s}) = \mu$$

$$\mathrm{var}(y_t) = \mathrm{var}(y_{t-s}) = \sigma_y^2$$

$$\mathrm{cov}(y_t, y_{t-s}) = \mathrm{cov}(y_{t-j}, y_{t-j-s}) = \gamma_s$$

i.e. first and second moments are **constants** *independent of time of observation*

Stronger versions exist: strictly stationary

General univariate time series process

The most general univariate time series process we will work with in this section:

$$y_t = a_0 + \underbrace{a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p}}_{\text{Auto-Regressive of order p}} + \overset{\substack{\text{White Noise} \\ \downarrow}}{\varepsilon_t} + \underbrace{b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}}_{\text{Moving Average error of order } q}$$

General univariate time series process

The most general univariate time series process we will work with in this section:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

Using Lag operators:

$$y_t = a_0 + \left(a_1 L + \cdots + a_p L^p\right) y_t + \left(1 + b_1 L \ldots + b_q L^q\right) \varepsilon_t$$

$$\left(1 - a_1 L - \cdots - a_p L^p\right) y_t = a_0 + \left(1 + b_1 L \ldots + b_q L^q\right) \varepsilon_t$$

$$A(L) y_t = a_0 + B(L) \varepsilon_t$$

Where we use compact notation: $A(L)$ and $B(L)$ are called *lag polynomials*

General univariate time series process

The most general univariate time series process we will work with in this section:

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

Now we can use the work from before:

the inverse characteristic polynomial for y_t is $A(L)$

- **If all roots of $\boldsymbol{A(L)}$ are outside the unit circle, $\boldsymbol{y_t}$ is *stationary***
  - Then the process is called an Auto-Regressive, Moving Average process: ARMA(p,q)
  - This hold for any finite order $B(L)$. The roots of $B(L)$ determine if the process is ***invertible***

General univariate time series process

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

- If all roots of $A(L)$ are outside the unit circle, y_t is stationary
  - Then the process is called an Auto-Regressive, Moving Average process: $ARMA(p, q)$
  - This hold for any finite order $B(L)$. The roots of $B(L)$ determine if the process is **invertible**

- If one or more roots of $A(L)$ are equal to 1, and the rest outside the unit circle y_t is called *integrated*, and is *non-stationary*.
  - Behaves like a random walk, wanders or drifts with no reversion to a constant mean, moreover, the unconditional variance is undefined
  - It is called an Auto-Regressive, Integrated, Moving Average process: $ARIMA(p, d, q)$
  - More on this later
  - An $ARIMA(p, d, q)$ process is stationary if it is differenced d times.

Representation results (briefly)

Consider the stationary ARMA(p,q) process:

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

Since y_t is stationary, all roots of $A(L) = 0$ are outside the unit circle

$\Rightarrow A(L)^{-1}$ is well defined

$\Rightarrow y_t = A(L)^{-1}(a_0 + B(L)\varepsilon_t)$ is an MA(∞) process

Recall from before, if $|a_1| < 1$:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \Leftrightarrow y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

Representation results (briefly)

Consider the stationary ARMA(p,q) process:

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

If all roots of $B(L) = 0$ are outside the unit circle

$\Rightarrow B(L)^{-1}$ is well defined

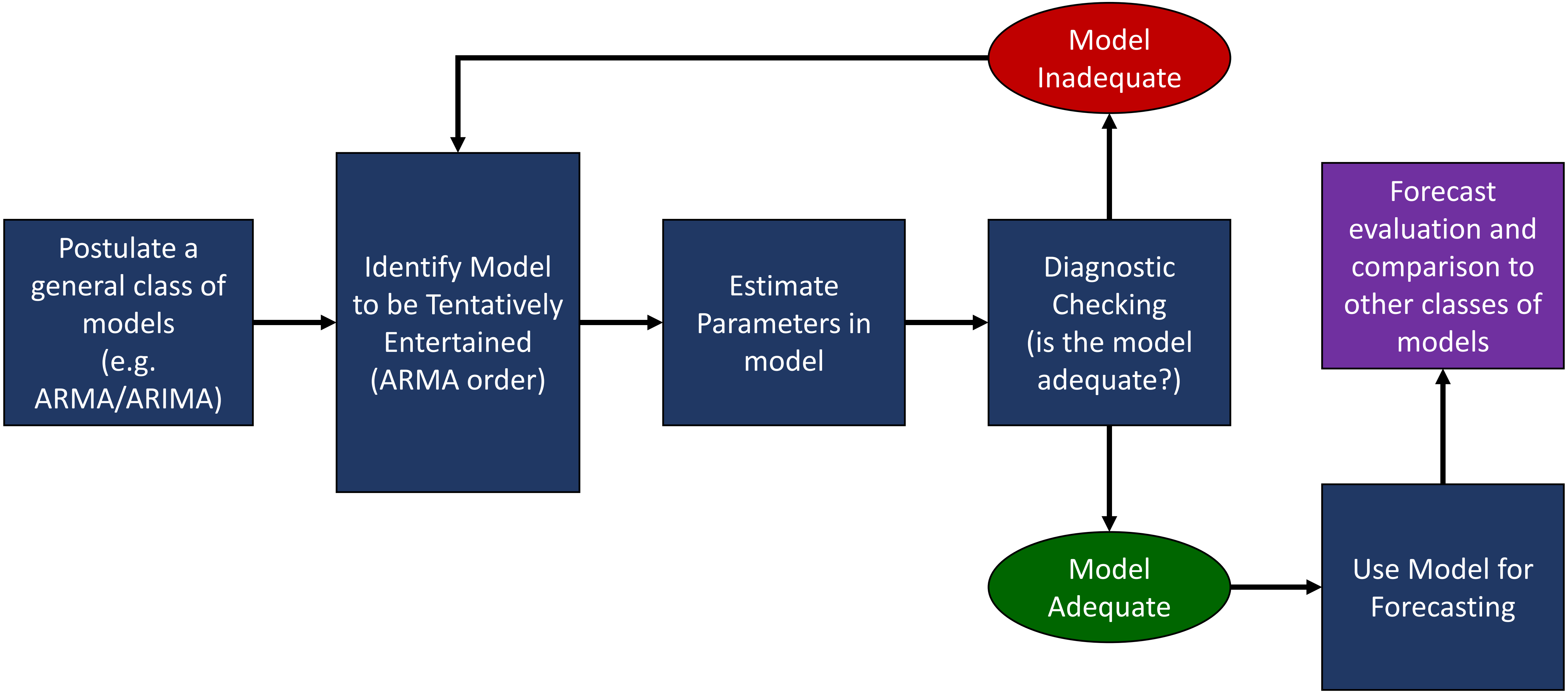
$$\Rightarrow \varepsilon_t = B(L)^{-1}(-a_0 + A(L)y_t)$$
$$= \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \cdots$$, is a convergent AR(∞) process

We will not use this result much, but it is a requirement for using the Box-Jenkins method to analyse univariate processes.

Univariate modelling for forecasts

- We will be applying the Box-Jenkins method
    - Originally developed for forecasting/control of management, economic and physical processes
    - In principal, it is problematic to apply a simple stationary, univariate ARMA model to the highly endogenous/interrelated nature of economic time series
    - Still a prevalent method for a simple benchmark to compare forecasting performance of more detailed models
    - It turns out to be very difficult to do better than forecasting with a simple ARMA process

The Box-Jenkins approach



Adapted from Box and Jenkins (1991)

Plan for the day/week

- Work through finding the moments of different ARMA processes
  - This will lead us to what the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are
- Discuss the patterns expected from ARMA(p,q) models, depending on the nature of the lag polynomials that characterize the different parts
- Consider the estimation and diagnostic tests of models
- Define forecasts and their evaluation
- Simulate, estimate and forecast ARMA processes

Moments of stationary ARMA processes

- We will consider a few simple examples by hand – AR(1), MA(q) – as these will establish the basic ideas
- We will find the following moments of these processes

$$\mathbb{E}(y_t)$$

$$\mathrm{var}(y_t) = \mathbb{E}\big(y_t - \mathbb{E}(y_t)\big)^2$$

$$\mathrm{cov}(y_t, y_{t-s}) = \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-s} - \mathbb{E}(y_{t-s})\big)\big]$$

$$\mathrm{cor}(y_t, y_{t-s}) = \frac{\mathrm{cov}(y_t, y_{t-s})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_{t-s})}}$$

Moments of stationary ARMA processes

- Things quickly get messy for AR(2) and above – Enders and Hamilton give strong guides to the methodologies for an arbitrary ARMA process
  - These are very useful to build your toolkit, and often used in the solution of macro models, but not particularly relevant for empirical work
  - One can get a sufficiently strong intuition via simulation, which we will do in Matlab
  - Stationarity is the central concept which we have already established the rule for
- This is often presented as a "check for stationarity":
  - find the moments, if they are constants independent of time, conclude stationarity. We will skip this an assume we are working with stationary processes, as we already have the rule

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- The unconditional expectation of y_t
- Formal definition?

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- The unconditional expectation of y_t
- Formal definition?
  - Probability weighted average of all possible values that y_t may take

- Let $f(y_t)$ be the density function of the random variable y_t, then

$$\mathbb{E}(y_t) = \int_{-\infty}^{\infty} y_t \, f(y_t) dy_t$$

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- Empirically, we observe a time series for, say GDP:

$$\{\dots, y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, \dots\}$$

- One's "cross section intuition" might suggest that the expected value of y_t is the limit of the time-mean (by the law of large numbers)

$$"\mathbb{E}(y_t) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} y_t \text{ “}$$

- In fundamental terms, however, this is not *always* true

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- Empirically, we observe a time series for, say GDP:

$$\{\dots, y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, \dots\}$$

- Since we know there is dependence over time, this counts as **only ONE** observation of the dynamic Data Generating Process (DGP)

$$y_t = f(y_{t-1}, y_{t-2}, \dots)$$

- Thus, in general

$$\mathbb{E}(y_t) \not\equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} y_t$$

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- Philosophically, consider an large number of runs of the economy (or abstractly, a large number of draws of time *paths* from the same Data Generating Process): These are called an *ensemble* of different observations of the process over time

$$\left\{\dots, y_{t-2}^{(i-1)}, y_{t-1}^{(i-1)}, y_{t}^{(i-1)}, y_{t+1}^{(i-1)}, y_{t+2}^{(i-1)}, \dots\right\}$$

$$\left\{\dots, y_{t-2}^{(i)}, y_{t-1}^{(i)}, y_{t}^{(i)}, y_{t+1}^{(i)}, y_{t+2}^{(i)}, \dots\right\}$$

$$\left\{\dots, y_{t-2}^{(i+1)}, y_{t-1}^{(i+1)}, y_{t}^{(i+1)}, y_{t+1}^{(i+1)}, y_{t+2}^{(i+1)}, \dots\right\}$$

- **If it exists**, the Law of Large numbers says that $\mathbb{E}(y_t)$ is equal to the limit of the *ensemble mean* of the value of the process y at time t:

$$\mathbb{E}(y_t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} y_t^{(i)}$$

Technical considerations

what do we mean by $\mathbb{E}(y_t)$?

- A data generating process is called ***ergodic*** if

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T} y_t = \lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N} y_t^{(i)}$$

- This is why the theory of time series econometrics starts with stationary processes:
    - Stationary processes are ergodic (under some additional regularity conditions), so we can use the limit of time averages as the moments of the process
    - Non stationary processes are not ergodic, hence we need different approaches
    - We will, for instance, “derive” the critical values for the Dickey Fuller unit root tests by Monte Carlo simulation (which is what the authors also did)

White Noise

- As always in this course, ε_t denotes white noise:
- The formal characteristics of which are?

White Noise

- As always in this course, ε_t denotes white noise:

$$\mathbb{E}(\varepsilon_t) = 0$$

$$\mathbb{E}(\varepsilon_t^2) = \sigma^2$$

$$\mathbb{E}(\varepsilon_t \varepsilon_s) = 0 \;\; \forall \;\; t \neq s$$

Independent of any other stochastic process

Finding the moments of ARMA processes

- For the next few slides, we will be finding the moments of various ARMA(p,q) processes, assuming stationarity

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$

Next, we use the fact that the expectation operator is *linear*:

The expected value of a sum of random variables

is equal to

the sum of the expected values of those random variables

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$
$$= \mu + \mathbb{E}(\varepsilon_t) + b\mathbb{E}(\varepsilon_{t-1})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathbb{E}(y_t)$:

$$\begin{aligned} \mathbb{E}(y_t) &= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1}) \\ &= \mu + \mathbb{E}(\varepsilon_t) + b\mathbb{E}(\varepsilon_{t-1}) \\ &= \mu \end{aligned}$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{var}(y_t)$:

$$\text{var}(y_t) = \text{var}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$

- Rule for variance: a, b and c constants, X and Y random variables

$$\text{var}(a + bX + cY) =$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{var}(y_t)$:

$$\mathrm{var}(y_t) = \mathrm{var}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$

- Rule for variance: a, b and c constants, X and Y random variables

$$\mathrm{var}(a + bX + cY) = b^2\mathrm{var}(X) + c^2\mathrm{var}(Y) + 2bc \cdot \mathrm{cov}(X, Y)$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{var}(y_t)$:

$$\mathrm{var}(y_t) = \mathrm{var}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{var}(y_t)$:

$$\mathrm{var}(y_t) = \mathrm{var}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$
$$= \mathrm{var}(\varepsilon_t) + b^2\mathrm{var}(\varepsilon_{t-1}) + 2b \cdot \mathrm{cov}(\varepsilon_t, \varepsilon_{t-1})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{var}(y_t)$:

$$\text{var}(y_t) = \text{var}(\mu + \varepsilon_t + b\varepsilon_{t-1})$$
$$= \text{var}(\varepsilon_t) + b^2\text{var}(\varepsilon_{t-1}) + 2b \cdot \text{cov}(\varepsilon_t, \varepsilon_{t-1})$$
$$= \sigma^2 + b^2\sigma^2 + 0$$
$$= (1 + b^2)\sigma^2$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\operatorname{cov}(y_t, y_{t-1})$:

$$\operatorname{cov}(y_t, y_{t-1}) = \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-1} - \mathbb{E}(y_{t-1})\big)\big]$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{cov}(y_t, y_{t-1})$:

$$\text{cov}(y_t, y_{t-1}) = \mathbb{E}\left[\left(y_t - \mathbb{E}(y_t)\right)\left(y_{t-1} - \mathbb{E}(y_{t-1})\right)\right]$$
$$= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-1} + b\varepsilon_{t-2} - \mu)$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{cov}(y_t, y_{t-1})$:

$$\text{cov}(y_t, y_{t-1}) = \mathbb{E}\left[\left(y_t - \mathbb{E}(y_t)\right)\left(y_{t-1} - \mathbb{E}(y_{t-1})\right)\right]$$
$$= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-1} + b\varepsilon_{t-2} - \mu)$$
$$= \mathbb{E}(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t-1} + b\varepsilon_{t-2})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{cov}(y_t, y_{t-1})$:

$$\text{cov}(y_t, y_{t-1}) = \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-1} - \mathbb{E}(y_{t-1})\big)\big]$$
$$= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-1} + b\varepsilon_{t-2} - \mu)$$
$$= \mathbb{E}(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t-1} + b\varepsilon_{t-2})$$
$$= \mathbb{E}(\varepsilon_t\varepsilon_{t-1} + b\varepsilon_t\varepsilon_{t-2} + b\varepsilon_{t-1}\varepsilon_{t-1} + b^2\varepsilon_{t-1}\varepsilon_{t-2})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{cov}(y_t, y_{t-1})$:

$$\text{cov}(y_t, y_{t-1}) = \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-1} - \mathbb{E}(y_{t-1})\big)\big]$$
$$= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-1} + b\varepsilon_{t-2} - \mu)$$
$$= \mathbb{E}(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t-1} + b\varepsilon_{t-2})$$
$$= \mathbb{E}(\varepsilon_t\varepsilon_{t-1} + b\varepsilon_t\varepsilon_{t-2} + b\varepsilon_{t-1}\varepsilon_{t-1} + b^2\varepsilon_{t-1}\varepsilon_{t-2})$$
$$= \mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + b\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + b\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-1}) + b^2\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-2})$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\text{cov}(y_t, y_{t-1})$:

$$\begin{aligned}
\text{cov}(y_t, y_{t-1}) &= \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-1} - \mathbb{E}(y_{t-1})\big)\big] \\
&= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-1} + b\varepsilon_{t-2} - \mu) \\
&= \mathbb{E}(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t-1} + b\varepsilon_{t-2}) \\
&= \mathbb{E}(\varepsilon_t\varepsilon_{t-1} + b\varepsilon_t\varepsilon_{t-2} + b\varepsilon_{t-1}\varepsilon_{t-1} + b^2\varepsilon_{t-1}\varepsilon_{t-2}) \\
&= \mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + b\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) + b\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-1}) + b^2\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-2}) \\
&= b\sigma^2
\end{aligned}$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{cov}(y_t, y_{t-2})$:

$$\begin{aligned}
&\mathrm{cov}(y_t, y_{t-2}) = \mathbb{E}\big[\big(y_t - \mathbb{E}(y_t)\big)\big(y_{t-2} - \mathbb{E}(y_{t-2})\big)\big] \\
&= \mathbb{E}(\mu + \varepsilon_t + b\varepsilon_{t-1} - \mu)(\mu + \varepsilon_{t-2} + b\varepsilon_{t-3} - \mu) \\
&= \mathbb{E}(\varepsilon_t + b\varepsilon_{t-1})(\varepsilon_{t-2} + b\varepsilon_{t-3}) \\
&= \mathbb{E}(\varepsilon_t\varepsilon_{t-2} + b\varepsilon_t\varepsilon_{t-3} + b\varepsilon_{t-1}\varepsilon_{t-2} + b^2\varepsilon_{t-1}\varepsilon_{t-3}) \\
&= \mathbb{E}(\varepsilon_t\varepsilon_{t-2}) + b\mathbb{E}(\varepsilon_t\varepsilon_{t-3}) + b\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-2}) + b^2\mathbb{E}(\varepsilon_{t-1}\varepsilon_{t-3}) \\
&= 0
\end{aligned}$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{cor}(y_t, y_{t-1}) = \frac{\mathrm{cov}(y_t, y_{t-1})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_{t-1})}} = \frac{\mathrm{cov}(y_t, y_{t-1})}{\mathrm{var}(y_t)}$:

$$\mathrm{cor}(y_t, y_{t-1}) = \frac{b\sigma^2}{(1+b^2)\sigma^2} = \frac{b}{(1+b^2)}$$

Moments of stationary MA(1) process

$$y_t = \mu + \varepsilon_t + b\varepsilon_{t-1}$$

- $\mathrm{cor}(y_t, y_{t-1}) = \frac{\mathrm{cov}(y_t, y_{t-1})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_{t-1})}} = \frac{\mathrm{cov}(y_t, y_{t-1})}{\mathrm{var}(y_t)}$:

$$\gamma_0 = \mathrm{cor}(y_t, y_t) = \frac{(1+b^2)\sigma^2}{(1+b^2)\sigma^2} = 1$$

$$\gamma_1 = \mathrm{cor}(y_t, y_{t-1}) = \frac{b\sigma^2}{(1+b^2)\sigma^2} = \frac{b}{(1+b^2)}$$

$$\gamma_s = \mathrm{cor}(y_t, y_{t-s}) = 0 \ \forall\ s > 1$$

- The sequence: $\{\gamma_0, \gamma_1, \gamma_2, \dots\}$ is called the **Autocorrelation Function** of the process

Homework

- Find the autocorrelation functions of:

$$y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

$$y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

$$y_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(a_0 + a_1 y_{t-1} + \varepsilon_t)$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(a_0 + a_1 y_{t-1} + \varepsilon_t)$$
$$= a_0 + a_1 \mathbb{E}(y_{t-1}) + \mathbb{E}(\varepsilon_t)$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(a_0 + a_1 y_{t-1} + \varepsilon_t)$$
$$= a_0 + a_1 \mathbb{E}(y_{t-1}) + \mathbb{E}(\varepsilon_t)$$
$$= a_0 + a_1 \mathbb{E}(y_t) + 0$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}(a_0 + a_1 y_{t-1} + \varepsilon_t)$$
$$= a_0 + a_1 \mathbb{E}(y_{t-1}) + \mathbb{E}(\varepsilon_t)$$
$$= a_0 + a_1 \mathbb{E}(y_t) + 0$$
$$= \frac{a_0}{1 - a_1}$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \, \varepsilon_{t-i}$$

- $\mathbb{E}(y_t)$:

$$\mathbb{E}(y_t) = \mathbb{E}\left(\frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \, \varepsilon_{t-i}\right)$$

$$= \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \, \mathbb{E}(\varepsilon_{t-i})$$

$$= \frac{a_0}{1 - a_1}$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathrm{var}(y_t)$:
- It is instructive to use the backward solution:

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \, \varepsilon_{t-i}$$

- And to note:

$$y_t - \mathbb{E}(y_t) = \sum_{i=0}^{\infty} a_1^i \, \varepsilon_{t-i} \ \forall \ t$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\text{var}(y_t)$:

$$\text{var}(y_t) = \mathbb{E}\big(y_t - \mathbb{E}(y_t)\big)^2$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathrm{var}(y_t)$:

$$\mathrm{var}(y_t) = \mathbb{E}\big(y_t - \mathbb{E}(y_t)\big)^2$$

$$= \mathbb{E}([\varepsilon_t + a_1 \varepsilon_{t-1} + a_1^2 \varepsilon_{t-2} + \cdots][\varepsilon_t + a_1 \varepsilon_{t-1} + a_1^2 \varepsilon_{t-2} + \cdots])$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\mathrm{var}(y_t)$:

$$\mathrm{var}(y_t) = \mathbb{E}\big(y_t - \mathbb{E}(y_t)\big)^2$$
$$= \mathbb{E}([\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots][\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots])$$
$$= \sigma^2 + a_1^2\sigma^2 + a_1^4\sigma^2 + \cdots$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\text{var}(y_t)$:

$$\text{var}(y_t) = \mathbb{E}\big(y_t - \mathbb{E}(y_t)\big)^2$$
$$= \mathbb{E}([\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots][\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots])$$
$$= \sigma^2 + a_1^2\sigma^2 + a_1^4\sigma^2 + \cdots$$
$$= \sigma^2[1 + a_1^2 + a_1^4 + a_1^6 + \cdots]$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- $\text{cov}(y_t, y_{t-s})$:

$$\text{cov}(y_t, y_{t-s}) = \mathbb{E}\left[\left(y_t - \mathbb{E}(y_t)\right)\left(y_{t-s} - \mathbb{E}(y_{t-s})\right)\right]$$
$$= \mathbb{E}([\varepsilon_t + \cdots + a_1^s \varepsilon_{t-s} + a_1^{s+1}\varepsilon_{t-s-1} + \cdots][\varepsilon_{t-s} + a_1 \varepsilon_{t-s-1} + a_1^2 \varepsilon_{t-s-2} + \cdots])$$
$$= a_1^s \sigma^2 + a_1^{s+2}\sigma^2 + a_1^{s+4}\sigma^2 + \cdots$$
$$= a_1^s \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \cdots]$$
$$= \frac{a_1^s \sigma^2}{1 - a_1^2}$$

Moments of stationary AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- Autocorrelation function:

$$\gamma_0 = \mathrm{cor}(y_t, y_t) = 1$$

$$\gamma_s = \frac{\mathrm{cov}(y_t, y_{t-s})}{\sqrt{\mathrm{var}(y_t)}\sqrt{\mathrm{var}(y_{t-s})}} = \frac{\mathrm{cov}(y_t, y_{t-s})}{\mathrm{var}(y_t)} = \frac{\dfrac{a_1^s \sigma^2}{1 - a_1^2}}{\dfrac{\sigma^2}{1 - a_1^2}} = a_1^s$$

Moments of stationary AR(1) process

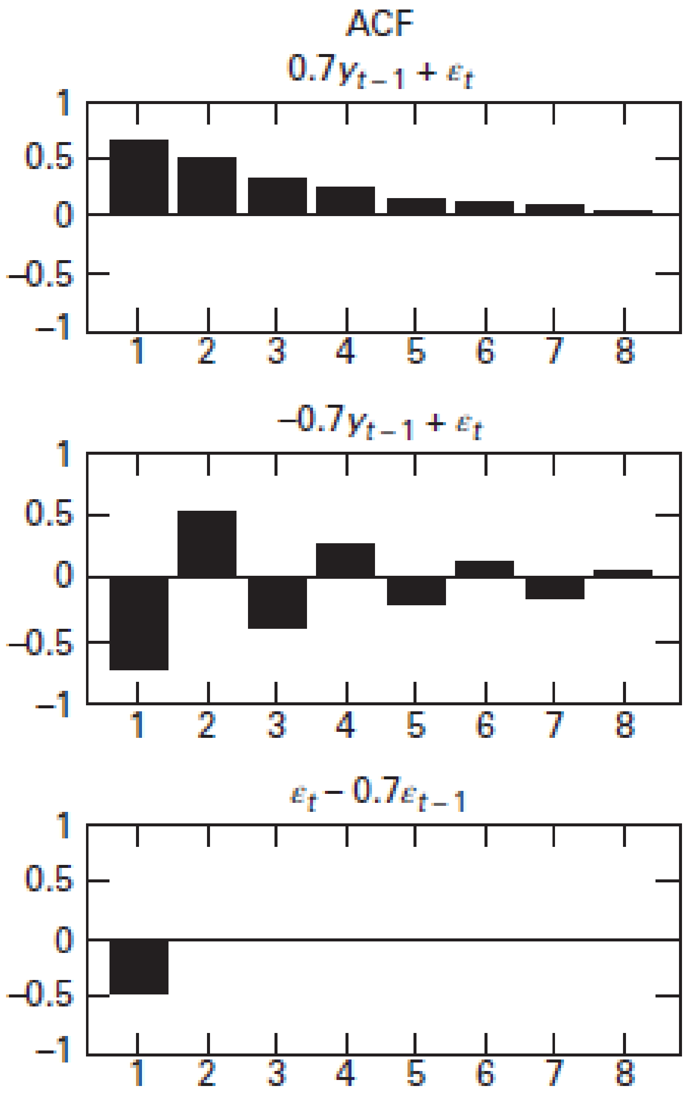
$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- Autocorrelation function:

$$\{1, a_1, a_1^2, a_1^3, \dots\}$$

- Since we assumed stationarity, we know $|a_1| < 1$
  - If $0 < a_1 < 1$, the ACF will be a positive and geometrically decay
  - If $-1 < a_1 < 0$, the ACF will alternate between positive and negative values, but still geometrically decay

ACF patterns of stationary ARMA(p,q) processes



Partial Autocorrelation Function

- The ACF describes the “unconditional” correlation between y_t and y_{t-s}
- The PACF considers the correlation between y_t and y_{t-s}, conditional on the values between the two dates
- The PACF can be derived analytically, but it is sufficient to just consider the “regression based” definition to get the intuition

Partial Autocorrelation Function

- For some time series $\{y_t\}$, consider the following sequence of OLS regressions:

$$
\begin{aligned}
&y_t = \phi_{01} + \phi_{11} y_{t-1} + u_t \\
&y_t = \phi_{02} + \phi_{12} y_{t-1} + \phi_{22} y_{t-2} + u_t \\
&y_t = \phi_{03} + \phi_{13} y_{t-1} + \phi_{23} y_{t-2} + \phi_{33} y_{t-3} + u_t \\
&\vdots
\end{aligned}
$$

- The PACF is defined as the sequence $\{1, \phi_{11}, \phi_{22}, \phi_{33}, \dots\}$
  - Since OLS estimators have closed form solutions, we can, in principle, find analytical versions of these concepts for any given ARMA process, known as the *theoretical PACF*, which can then be compared to the estimated versions in the Box-Jenkins methodology. Hamilton and Box and Jenkins give these analytical exercises.

Theoretical
ACF and PACF patterns
for ARMA processes

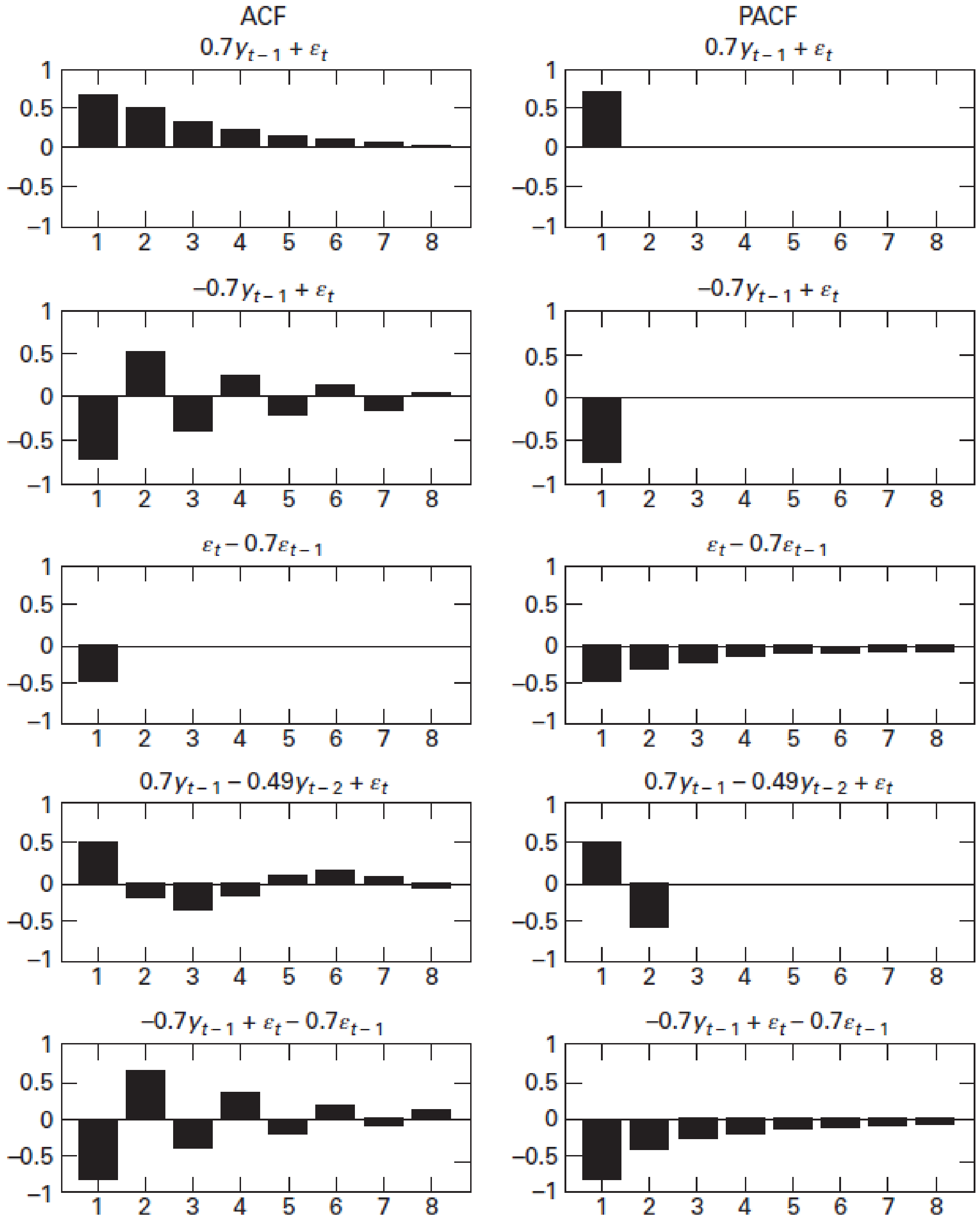


FIGURE 2.2 Theoretical ACF and PACF Patterns

Table 2.1 Properties of the ACF and PACF

Process	**ACF**	**PACF**
White noise	All $\rho_s = 0$ $(s \neq 0)$	All $\phi_{ss} = 0$
AR(1): $a_1 > 0$	Direct geometric decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p. All $\phi_{ss} = 0$ for $s > p$
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Oscillating decay: $\phi_{11} > 0$
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Geometric decay: $\phi_{11} < 0$
ARMA(1, 1): $a_1 > 0$	Geometric decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1): $a_1 < 0$	Oscillating decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q	Decay (either direct or oscillatory) beginning after lag p

Plan for the week

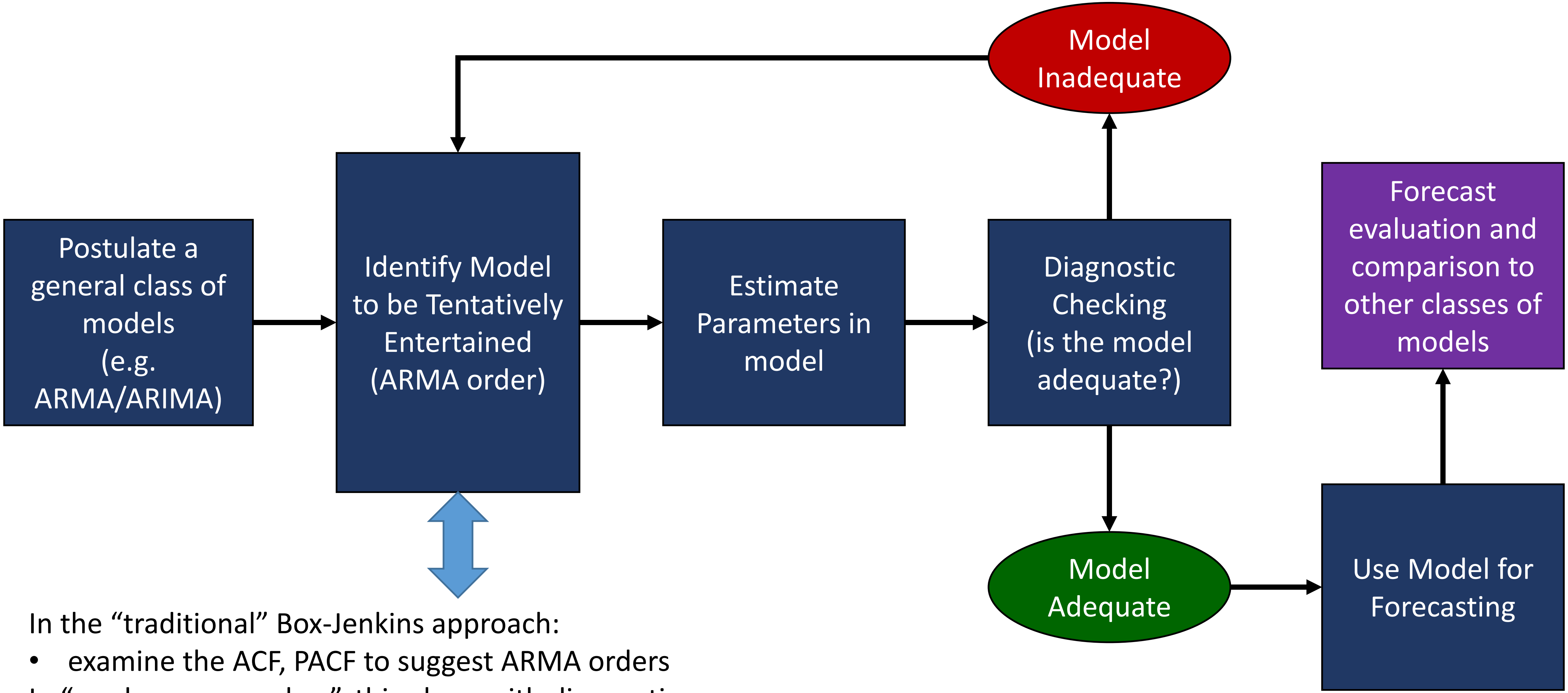
Above:

- Defined stationarity and worked through finding the moments of different stationary ARMA processes
  - This lead us to the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF)
- Discussed the patterns expected from ARMA(p,q) models, depending on the nature of the lag polynomials that characterize the different parts

Now we use this:

- Consider the estimation and diagnostic tests of models
- Define forecasts and their evaluation
- Simulate, estimate and forecast ARMA processes in Matlab

The Box-Jenkins approach



In the “traditional” Box-Jenkins approach:

- examine the ACF, PACF to suggest ARMA orders

In “modern approaches”, this along with diagnostic checking is typically automated, but this is risky

Adapted from Box and Jenkins (1991)

Estimation

- AR(p) models are linear in parameters with known "regressors" and can be estimated by OLS. The errors in the equation are simply estimated as the OLS residuals
- MA(q) and ARMA(p,q) models depend on past values of errors, which themselves are part of the realization of the process in several periods, hence cannot be reduced to OLS estimators with closed form solutions
  - These are estimated via Maximum Likelihood
  - We will not delve into this today, as these have been coded in any good package

Diagnostic Checking/Model Selection

- There are two key requirements of an adequate model of a time series:

1. It must be **congruent** with the Data Generating process
2. It must be **parsimonious**

- This is how David Hendry presents his view, and I have not found a better formal approach
  - It’s easy to get spurious or misleading results without such a systematic approach

Diagnostic Checking/Model Selection

- There are two key requirements of an adequate model of a time series:

1. It must be **congruent** with the Data Generating process

- This means the model must capture all the time series properties of the data.
  - Put differently, the empirical model must not have features not postulated in the population model
- A minimum requirement for this is that the *residuals* from the *estimated* model are white noise
  - I.e. the unexplained part of the model must contain no systematic time series information

Diagnostic Checking/Model Selection

1. A model must be **congruent** with the Data Generating process

- This means the model must capture all the time series properties of the data.
- A minimum requirement for this is that the residuals from the model are white noise
- This poses a philosophical problem with allowing a “moving average process in errors”. What does that mean as a description of an economic problem?
  - In the univariate context, there is not much we can say about this, and we will treat it mechanically – there are ARMA(p,q) processes, and we will estimate them as such
  - In the general abstract forecasting literature, it is also less notable: the goal is finding the model that forecasts the best, not to identify causal effects. Whether it matches the a sensible economic model or not is less important. However, I have suspicions on the likely new sample performance of relying on MA terms
  - When we wish to make fundamental statements about the true nature of the underlying economic process, this is deeply problematic: The residuals in an estimation are not the “errors of the economic process”, they are the parts of the dynamics we cannot explain. Arbitrarily giving “Structure to what we do not understand so that the model fits” means (to me) that the model is junk

Diagnostic Checking/Model Selection

1. A model must be **congruent** the Data Generating process

- This means the model must capture all the time series properties of the data.
- A minimum requirement for this is that the residuals from the model are white noise
- When choosing among a set of models, one can consider “the relative degree” to which different models encompass the data: i.e. which model explains more of the variation?
- We have the typical measures: which model has produces the smallest sum of squared errors/highest likelihood? (there are many different versions depending on the model type)
  - As usual, there is a trade-off: adding more coefficients necessarily reduces the sum of squared errors, but may not make the model better, hence the other requirement

Diagnostic Checking/Model Selection

2. A model must be **parsimonious**

- A larger model (more parameters to estimate) will always fit better *within the estimation sample*
- However, the more parameters to estimate, the fewer the available data points that are available to identify the value of each parameter, and hence the worse the accuracy of point estimate of each parameter
- The lower the accuracy of a parameter estimate, the worse the forecast performance, or in other words, the worse the *out of sample* performance
- A minimum requirement of a parsimonious model is that all coefficients should be statistically significant (although there are nuances to this rule in multivariate context)

Diagnostic Checking/Model Selection

2. A model must be **parsimonious**

- Within the ARMA context, there is another specific issue that makes the model selection difficult
- Representation results: a stationary AR(p) model has a MA(∞) representation, and an invertible MA(q) process has a AR(∞) representation. Thus data generated by an AR(1) process might be fit well by and MA(2) or (3) model, but the AR(1) is more parsimonious
- Given data from a ARMA(p,q) of known order, the best empirical model of the data is *not necessarily* the one that imposes the actual order of the process – a more parsimonious model does better
- Common factors:

$$y_t = 0.25y_{t-2} + \varepsilon_t + 0.5\varepsilon_{t-1}$$

$$(1 - 0.25L^2)y_t = (1 + 0.5L)\varepsilon_t$$

Diagnostic Checking/Model Selection

2. A model must be **parsimonious**

- Within the ARMA context, there is another specific issue that makes the model selection difficult
- Representation results: a stationary AR(p) model has an MA(∞) representation, and an invertible MA(q) process has an AR(∞) representation. Thus data generated by an AR(1) process might be fit well by and MA(2) or (3) model, but the AR(1) is more parsimonious
- Given data from a ARMA(p,q) of known order, the best empirical model of the data is *not necessarily* the one that imposes the actual order of the process – a more parsimonious model does better
- Common factors:

$$(1-0.25L^2)y_t = (1+0.5L)\varepsilon_t$$

$$(1+0.5L)(1-0.5L)y_t = (1+0.5L)\varepsilon_t$$

$$(1-0.5L)y_t = \varepsilon_t$$

$$y_t = 0.5y_{t-1} + \varepsilon_t$$

Diagnostic Checking/Model Selection

2. A model must be **parsimonious**

- Within the ARMA context, there is another specific issue that makes the model selection difficult
- Representation results: a stationary AR(p) model has a MA(∞) representation, and an invertible MA(q) process has a AR(∞) representation. Thus data generated by an AR(1) process might be fit well by and MA(2) or (3) model, but the AR(1) is more parsimonious
- Given data from a ARMA(p,q) of known order, the best empirical model of the data is *not necessarily* the one that imposes the actual order of the process – a more parsimonious model does better
- Common factors
- Enders also warns that highly correlated (close to collinear) coefficients are unstable and should be tested for elimination

Diagnostic Tests

White noise residuals:

- Plot of residuals
  - no more than 5% of residuals should be outside 95% confidence bands (normality/outliers)
  - Clusters of large residuals may indicate structural change (or ARCH behaviour)
- ACF of residuals
  - In a large enough sample, some will exceed 95%, but should not be systematic
- Q statistics: Tests groups of auto-correlations for joint significance
  - Box-Pierce: $\mathrm{Q} = T\sum_{k=1}^{s} r_k \sim \chi^2(s-n)$ where r_k is the k^{th} autocorrelation. Poor in large sample
  - Ljung-Box: $\mathrm{Q} = T(T+2)\sum_{k=1}^{s} \frac{r_k}{T-k} \sim \chi^2(s-n)$ (n is number of estimated parameters)

Selection Criteria

Two common measures that trade off better fit and parsimony:

- As R reports them = smaller is better
- Akaike Information Criterion (AIC) $= -\frac{2 \ln L}{T} + \frac{2n}{T}$
- Schwartz Bayesian Criterion (SBC) $= -\frac{2 \ln L}{T} + \frac{n \ln T}{T}$
- Both tend to select overparameterized models, but the SBC is asymptotically consistent

Structural Breaks/Parameter instability

- Recall that our assumptions are that the *same* process generates data over the entire sample. I.e. the coefficients are constants

Chow breakpoint test

- Tests whether the model is significantly different before or after some t
  - Estimate the model in two subsamples, do an F test for difference in Sum of Squared Residuals

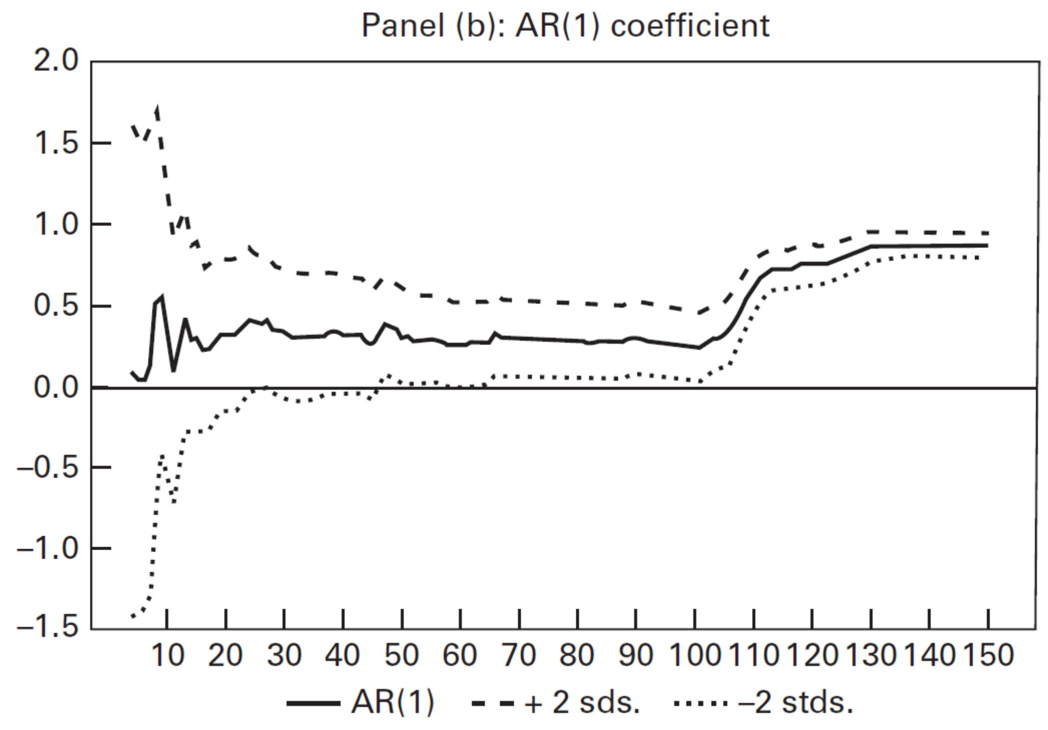
$$F = \frac{(\mathrm{SSR} - \mathrm{SSR}_1 - \mathrm{SSR}_2)/n}{(\mathrm{SSR}_1 + \mathrm{SSR}_2)/(T - 2n)}$$

Structural Breaks/Parameter instability

- Recall that our assumptions are that the *same* process generates data over the entire sample. I.e. the coefficients are constants

Parameter Stability checks: Recursive estimation

- Start with a small sample (say T=10), estimate parameters
- Successively as sample points, and re-estimate parameters
- Plot coefficient estimates and error bands, check if coefficients change significantly



Forecasts

- Hamilton (Chapter 3) gives an exhaustive treatment of the theory of "optimal forecasts" with infinite data, then translates this to approximations with finite data
  - This is particularly important for processes with MA components: how do you model the errors?
- We will treat "forecasts" naively

Conditional expectations:

- Consider the AR (1) model:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- Then:

$$y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1}$$

- The expectation conditional on everything known in period t

$$\begin{aligned} E_t(y_{t+1}) &= E_t(a_0 + a_1 y_t + \varepsilon_{t+1}) \\ &= a_0 + a_1 y_t + E_t(\varepsilon_{t+1}) \\ &= a_0 + a_1 y_t \end{aligned}$$

- The difference is just the new error (or *innovation*):

$$y_{t+1} - E_t(y_{t+1}) = \varepsilon_{t+1}$$

- Similarly, we can construct the conditional expectation of two periods ahead:

$$\begin{aligned} E_t(y_{t+2}) &= a_0 + a_1 E_t(y_{t+1}) + E_t(\varepsilon_{t+2}) \\ &= a_0 + a_1(a_0 + a_1 y_t) \end{aligned}$$

Forecasts with an estimated model:

- Consider an estimated AR (1) model:

$$y_t = \hat{a}_0 + \hat{a}_0 y_{t-1} + u_t$$

- Then the one step ahead forecast is:

$$\hat{y}_{t+1,t} = \hat{a}_0 + \hat{a}_0 y_t$$

- And the two step ahead forecast is:

$$\hat{y}_{t+2,t} = \hat{a}_0 + \hat{a}_0 \hat{y}_{t+1,t}$$

- And one can continue with this process to *h* step ahead forecasts
- This is what we will do in Matlab
- We will deal with forecast evaluation in more detail in a future lecture