GRA4159 Trends, Cycles, and Signal Extraction from a Macroeconomic Perspective

Time Series Fundamentals

Leif Anders Thorsrud

BI Norwegian Business School

2026

Summary from Lecture 1

You should know

- The importance of NAS for understanding and explaining macroeconomic dynamics
- What GDP actually is and its components and other important macro variables, such as CPI and interest rates
- The important difference between nominal and real values
- That NAS also have some issues related to reporting lags, revisions, and welfare
- That NAS are accounting identities and not causal relationships (e.g., equations describing how the economy works)

This lecture

NAS statistics and other macro variables are time series. To work with them we need to understand time series fundamentals

Agenda

1. What is a time series?
2. Some very important tools, definitions, and concepts
   - Lag operators, difference equations and the autocorrelation function
   - White noise and moving averages
   - Conditional and unconditional expectations
   - Stationarity
3. Some basic time series processes (building blocks)
   - Moving average process (MA)
   - Autoregressive process (AR)
   - Random Walk (RW)

2. What is a time series?

- Most data in macroeconomics and finance can be described as time series: a set of repeated observations over time of the same variable, such as consumer prices, gross domestic product (GDP), stock prices, exchange rates, etc.
- Accordingly, to be able to understand macroeconomic fluctuations and financial markets we need to know time series.
- In terms of the notation, we will from now specify y_t, where the subscript t runs over time, i.e. $t = 1, 2, 3..., T$.

What is a time series?

- Time series data can have a wide range of frequencies. This has implications for what type of analysis we can do on the data.
- The sample available for a given time series varies greatly. Many results in statistics and econometrics depend on having many observations.
- Time series data behave very differently. The logarithm of a given time series y_t can be thought of as the sum of four (additive) components:

$$y_t = g_t + c_t + s_t + \varepsilon_t$$

  where g_t is a trend, c_t is a cycle, s_t is a seasonal component, and ε_t is noise.

What is a time series?

What is a time series?

Trend:

- Either deterministic or stochastic trends. (more on this later)

Seasonal component:

- Deterministic wave movements. (think natural sciences).
- In economics think e.g. Christmas shopping, weekends, holidays,...

Noise:

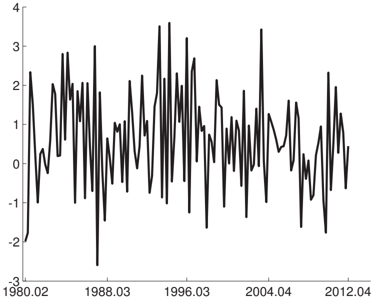
- Unpredictable/unanticipated innovation/shock.

What is the cycle?

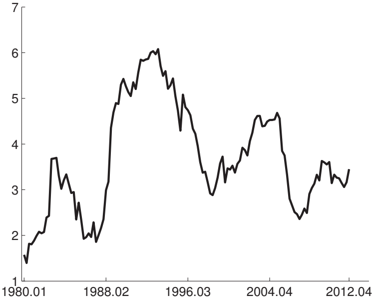
Cycle:

- Burns and Mitchell (1943):
  - expansions and recessions occurring at about the same time in many economic activities.
  - sequence of changes is **recurrent but not periodic**.
  - duration of business cycles vary from more than one year to ten or twelve years.
- Lucas (1977):
  - Movements about the trend in gross national product.
  - do **not** resemble deterministic wave motions.
  - cycles characterized by co-movement among different aggregative time series.
  - they are characterized by qualitative similarity and not by e.g. country or time period.

Various times series for Norway

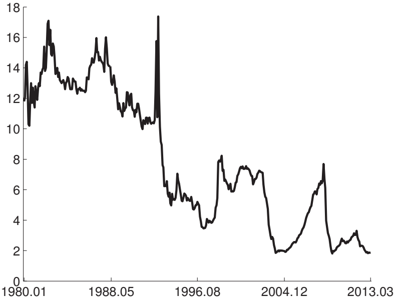


GDP growth (quarterly)



Unemployment rate (quarterly)

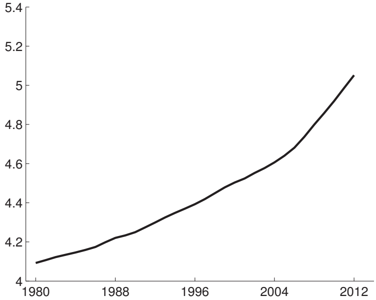
Various times series for Norway



3-month interest rate (monthly)

Oslo stock exchange (daily)

Various times series for Norway



Population (yearly)

Real house prices (yearly)

Our focus

- In this course we will focus mostly on the cyclical part of time series (+ noise), less on trends, and hardly nothing on the seasonal component
  - Most of the time we are not that interested in the seasonal component. However, estimating and removing the seasonal component from time series can be demanding and complicated and worth a course itself
  - Luckily, the statistical agencies typically do the job for us so that we can download seasonally adjusted NAS data directly from their databases
- Indeed, the theoretical model you will learn about in the following weeks will focus on **business cycles**, i.e., the cyclical part of time series. But, most of the time we need to estimate the trend to get the cycle. I.e.:
  - $\hat{c}_t = y_t + \hat{g}_t + \hat{\varepsilon}_t$ (when working with seasonally adjusted data)
  - Note also that in macro, depending on what variable y_t is, c_t is often refereed to as the output gap (for GDP), core inflation (for inflation), credit gap (for Credit), etc.
  - More on this later!

Defining time series

Defined formally, a time series is a collection of observations indexed by date of each observation, say starting in time $t = 1$ and ending in $t = T$:

$$\{y_1, y_2, y_3, \ldots, y_T\}$$

where the time index can be of any frequency, e.g., daily or quarterly. Usually this (finite) sample is a subset of an infinite sample, indexed by $\{y_t\}_{t=-\infty}^{\infty}$. A time series is usually identified by describing the t-th element.

We will generally treat y_t as a random variable, implying that a time series is a sequence of random variables ordered in time. We call such a sequence a stochastic process.

But, non-stochastic time series are also easy to envision (e.g. deterministic processes), i.e.

$$\begin{aligned} y_t &= t \\ y_t &= c \end{aligned} \tag{1}$$

Tools, definitions, and concepts

White noise, difference equations and lag operators

To introduce **randomness** into a time series process we work with white noise.

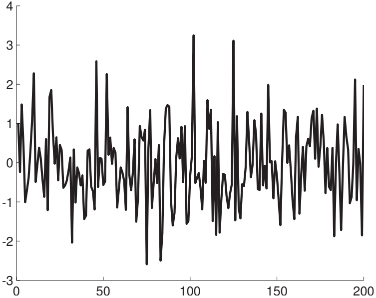
$$y_t = \varepsilon_t \tag{2}$$

where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a sequence of independent and random variables each of which has distribution:

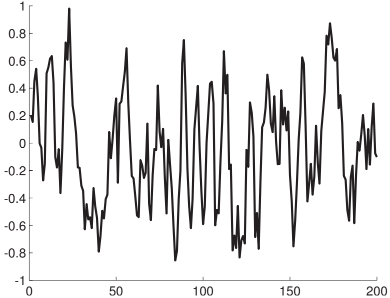
$$\varepsilon_t \sim iidN(0, \sigma^2) \tag{3}$$

That is, ε_t has mean zero and constant variance, σ^2.

White noise and moving average



White noise



Moving average

Moving average is here simply computed as $y_t = 1/5(\varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1} + \varepsilon_{t+2})$ which we call a centered moving average

Difference equations

To introduce **dynamics** we work with difference equations

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{4}$$

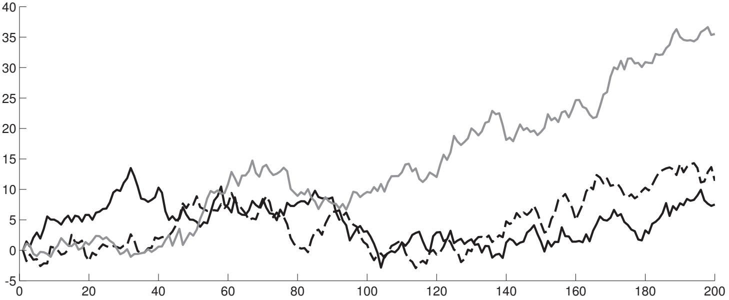
It is customary to assume that ε_t is white noise. If $\phi < 1$, we can show that the series will always return to its mean, i.e., it is (covariance) stationary.

If $\phi = 1$ we get a special, and important, case, called a random walk:

$$y_t = y_{t-1} + \varepsilon_t \tag{5}$$

More on this later

Different random walks



Lag operators

To find solutions to various difference equations it is helpful to employ lag operators, L, which transforms an observation at time t to period $t-1$, that is, backward one period in time:

$$Ly_t = y_{t-1} \tag{6}$$

Raising the lag operator to the power of -1 transforms the series one period forward, i.e., $L^{-1}y_t = y_{t+1}$ Generally, the lag operator can be raised to arbitrary integer powers k such that:

$$L^k y_t = y_{t-k}; L^{-k} y_t = y_{t+k}$$

A convenient use of the lag operator is to express the first difference of a series as:

$$\Delta y_t = (1-L)y_t = y_t - y_{t-1} \tag{7}$$

A polynomial of lag operators is called a lag polynomial, defined as $\phi(L)$, where:

$$\phi(L) = \left(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p\right) \tag{8}$$

where p is the lag order.

Expectations

We refer to the first moments of a stochastic process y_t as the mean:

$$\mu = \mathrm{E}\left[y_t\right], t = 1, ..., T$$

The first moments are interpreted as the average value of y_t taken over all possible realizations. Note that in this chapter we will not allow the mean of a time series to be time dependent, i.e., vary over time.

The second moments are defined as the variance:

$$\operatorname{var}\left(y_t\right) = \mathrm{E}\left[y_t, y_t\right] = \mathrm{E}\left[\left(y_t - E(y_t)\right)^2\right], t = 1, ..., T$$

and the covariance, for lag j:

$$\operatorname{cov}\left(y_t, y_{t-j}\right) = \mathrm{E}\left[y_t, y_{t-j}\right] = \mathrm{E}\left[\left(y_t - E(y_t)\right)\left(y_{t-j} - E(y_{t-j})\right)\right], t = j+1, ..., T$$

Conditional moments

The conditional distribution is based on the observation of some realization of random variables. Consider the following first-order difference equation:

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{9}$$

where $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2)$ and $|\phi| < 1$. Since the error terms are from the normal distribution, we know that y_t will also be a normally distributed variable. Conditional on information up to time t, i.e. knowing y_{t-1}, the mean, variance and covariance of the process in (9) are:

$$\begin{aligned}
\mathrm{E}\left[y_t \mid y_{t-1}\right] &= \phi y_{t-1} \\
\mathrm{var}\left(y_t \mid y_{t-1}\right) &= \sigma^2 \\
\mathrm{cov}\left(\left(y_t \mid y_{t-1}\right), \left(y_{t-j} \mid y_{t-j-1}\right)\right) &= 0 \text{ for} j > 1
\end{aligned} \tag{10}$$

This follows as ε_t is equal to 0 by construction, further that $\mathrm{var}\left(y_t \mid y_{t-1}\right) = \mathrm{E}\left[\phi y_{t-1} + \varepsilon_t - \phi y_{t-1}\right]^2 = \mathrm{E}\left[\varepsilon_t\right]^2 = \sigma^2$ and by employing the standard covariance formula. Hence, conditional on y_{t-1}, the distribution of y_t does not depend on previous values of y_{t-j}, for $j > 1$.

Conditional moments cont.

What happens if we instead condition on y_{t-2}?

$$\begin{aligned}
\mathrm{E}\left[y_t \mid y_{t-2}\right] &= \phi^2 y_{t-2} \\
\operatorname{var}\left(y_t \mid y_{t-2}\right) &= (1+\phi^2)\sigma^2 \\
\operatorname{cov}\left((y_t \mid y_{t-2}), (y_{t-j} \mid y_{t-j-2})\right) &= \phi\sigma^2 \text{ for } j = 1 \\
\operatorname{cov}\left((y_t \mid y_{t-2}), (y_{t-j} \mid y_{t-j-2})\right) &= 0 \text{ for } j > 1
\end{aligned} \tag{11}$$

Unconditional moments

Unconditional moments are the same as the populations moments, but explicitly specified in a time series context. The unconditional distribution of a time series process is arrived at under the hypothesis that no observation of the time series is available, i.e., we assume that we are at the observation preceding the realized observation. Thus, we only know the process generating the observation.

The unconditional moments of the process in (9) can be shown to be:

$$\begin{aligned} \mathrm{E}\left[y_t\right] &= 0 \\ \operatorname{var}\left(y_t\right) &= \frac{\sigma^2}{1-\phi^2} \\ \operatorname{cov}\left(y_t, y_{t-j}\right) &= \phi^j \operatorname{var}\left(y_t\right) \end{aligned} \tag{12}$$

Hence, in (10) the location of the mean depends on the conditioning information set, i.e., y_{t-1}. The unconditional mean in (12), on the other hand, is zero as we assume it only depends on the parameter ϕ. Also, knowing y_{t-1} changes the size of the variance of y_t and the degree of covariance with lagged values of y_t relative to the unconditional case.

What's the challenge?

In reality we only ever observe one realisation of a stochastic process, i.e. a sequence of T random variables. However, this one realisation is only one of infinitely many possible "paths of histories/alternative worlds".

We require time series properties that ensure that the amount of information contained in observing one realisation over time is equivalent to observing the process over and over again at the same point in time.

As we will see, the necessary properties are:

- stationarity
- ergodicity

What's the challenge?

Stationarity

- Stationarity is a fundamental concept in time series analysis.
- A time series is **strictly stationary** if for any values of $j_1, j_2, j_3, \ldots, j_n$, the joint distribution of $(y_t, y_{t+j_1}, y_{t+j_2}, y_{t+j_3}, \ldots, y_{t+j_n})$ depends only on the intervals separating the dates $(j_1, j_2, j_3, \ldots, j_n)$ and not on the date itself (t).
- If neither the mean (μ) nor the covariances $\mathrm{Cov}(y_t, y_{t-j})$ depend on the date t, then the process for y_t is said to be **covariance (weakly) stationary**:

$$\begin{aligned} E(y_t) &= \mu \\ E((y_t - \mu)(y_{t-j} - \mu)) &= Cov(y_t, y_{t-j}) \end{aligned} \qquad \text{for all } t, \text{ and any } j.$$

- Strong stationarity implies weak stationarity but not vice versa. Special case: Gaussian distribution.

Autocorrelation function (ACF)

When a process is stationary, its time domain properties can be summarized by computing the covariance of the process against a given number of lags. This is called the autocovariance function, and denoted $\gamma(j) \equiv cov(y_t, y_{t-1})$ for $t = 1, \ldots, T$ against j.

The autocovariance may be standardized by dividing by the variance of the process. This yields the autocorrelation function (ACF): $\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}$. Often it is useful to make a plot of $\rho(j)$ against (non-negative) j to learn about the properties of a given time series process. Note that $\rho(0) = 1$ by definition, as $\rho(0) \equiv \frac{\gamma(0)}{\gamma(0)} = 1$.

Ergodicity

A covariance-stationary process is said to be ergodic of the mean if the (sample) average $\bar{y} \equiv (1/T)\sum_{t=1}^{T} y_t$ converges in probability to $\mathrm{E}[y_t]$ as $T \to \infty$.

If ergodicity holds, it implies that the sample average and variance provides consistent estimates of their populations' counterparts. (Ergodic theorem).

One implication of ergodicity is that the autocorrelation function goes to zero quickly as j becomes large (observations sufficiently apart should be almost uncorrelated):

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty \tag{13}$$

That is, the dependencies between y_t and y_{t-j} weaken with increasing j. A stationary process is ergodic, if it is asymptotically independent, i.e. if any two random variables positioned far apart in the sequence are almost independently distributed.

- Can you think of a process that is stationary, but not ergodic?

Some basic time series processes (building blocks)

Moving average (MA) processes

We will now use the fundamentals discussed so far to construct and analyse the properties of time series models that make linear combinations of ε_t in different ways.

The first process we will consider in detail is the moving average (MA), which is a linear combination of white noise. The simplest case is the MA(1):

$$y_t = \delta + \varepsilon_t + \theta\varepsilon_{t-1} \tag{14}$$

where δ is a constant, y_t is the weighted sum of the two most recent values of ε, and ε_t is independent and identically distributed white noise, $\varepsilon_t \sim N(0, \sigma^2)$.

Is the MA stationary?

To examine this we need to calculate the different moments. We start with the mean, i.e., the first moment of the process:

$$\begin{aligned} \mathrm{E}[y_t] &= \mathrm{E}[\delta + \varepsilon_t + \theta\varepsilon_{t-1}] \\ &= \delta + \mathrm{E}[\varepsilon_t] + \theta\,\mathrm{E}[\varepsilon_{t-1}] \\ &= \delta \end{aligned} \tag{15}$$

Since the ε's are i.i.d., their expectations are zero. Hence, the mean is found to be δ, which is constant.

Turning to the second moments, we first compute the variance as:

$$\begin{aligned} \mathrm{var}(y_t) &= \mathrm{E}[y_t - \mathrm{E}[y_t]]^2 \\ &= \mathrm{E}[(\delta + \varepsilon_t + \theta\varepsilon_{t-1}) - \delta]^2 \\ &= \mathrm{E}[\varepsilon_t]^2 + 2\theta\,\mathrm{E}[\varepsilon_t\varepsilon_{t-1}] + \mathrm{E}[\theta\varepsilon_{t-1}]^2 \\ &= \sigma^2 + 0 + \theta^2\sigma^2 \\ &= (1+\theta^2)\sigma^2 \end{aligned} \tag{16}$$

Is the MA stationary cont.?

the first covariance as:

$$\begin{aligned}
\operatorname{cov}(y_t, y_{t-1}) &= \mathrm{E}\left[(y_t - \mathrm{E}[y_t])(y_{t-1} - \mathrm{E}[y_{t-1}])\right] \\
&= \mathrm{E}\left[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})\right] \\
&= \mathrm{E}\left[\varepsilon_t\varepsilon_{t-1}\right] + \mathrm{E}\left[\theta\varepsilon_{t-1}^2\right] + \mathrm{E}\left[\theta\varepsilon_t\varepsilon_{t-2}\right] + \mathrm{E}\left[\theta^2\varepsilon_{t-1}\varepsilon_{t-2}\right] \\
&= 0 + \theta\sigma^2 + 0 + 0 \\
&= \theta\sigma^2
\end{aligned} \tag{17}$$

and for lag order $j > 1$:

$$\begin{aligned}
\operatorname{cov}(y_t, y_{t-j}) &= \mathrm{E}\left[(y_t - \mathrm{E}[y_t])(y_{t-j} - \mathrm{E}[y_{t-j}])\right] \\
&= \mathrm{E}\left[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1})\right] \\
&= 0 \ \text{ for } \ j > 1
\end{aligned} \tag{18}$$

Neither the mean (equation (15)), the variance (equation (16)) nor the covariances (equations (17) and (18)) depend on time, the MA(1) process is covariance stationary. Moreover, the MA(1) process is stationary regardless of the value of θ.

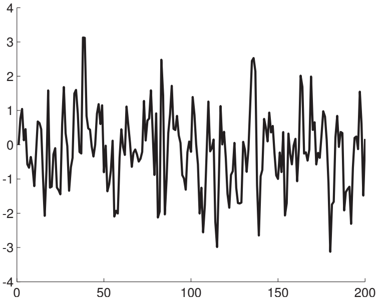
ACF for a MA process

What does the autocorrelation function (ACF) for an MA(1) process look like? Recall that the ACF is given by $\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}$. Accordingly, the autocorrelation for the MA(1) is:

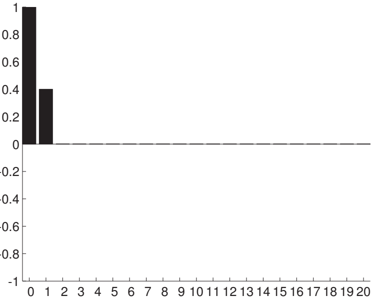
$$\begin{aligned} \rho(1) &= \frac{\theta}{(1+\theta^2)} \\ \rho(j) &= 0 \ \text{ for } j > 1 \end{aligned} \tag{19}$$

We see that for lag orders $j > 1$, the autocorrelations are zero. Hence, the autocorrelation function goes to zero as j becomes large, i.e., the MA(1) process is ergodic.

MA processes and ACF functions



MA $\theta = 0.5$



Autocorrelation function (ACF)

Summary thus far

1. The noise component is most often assumed to be Gaussian white noise.
2. Linear combinations of Gaussian white noise generate time series that behave similarly to many real world time series.
3. A time series is said to be covariance-stationary if neither its first nor second moments depend on the time index.
4. When a process is stationary, its time domain properties can be summarized by the autocorrelation function (ACF).
5. For a stationary process the ACF goes to zero quickly as the number of lags becomes larger.
6. The MA model is stationary regardless of its parameters.

Autoregressive (AR) processes

Another important time series process is the AR process:

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

The AR(1) relates the value of a variable y at time t, to its previous value at time $(t-1)$ and a random disturbance ε, also at time t. As always, ε_t is independent and identically distributed white noise, $\varepsilon_t \sim N(0, \sigma^2)$.

If $|\phi| < 1$, the AR(1) is covariance-stationary, with finite variance, i.e.:

$$\begin{aligned}
&\mathrm{E}[y_t] = 0 \\
&\mathrm{var}(y_t) = \tfrac{\sigma^2}{1-\phi} \\
&\mathrm{cov}(y_t, y_{t-j}) = \phi^j \,\mathrm{var}(y_t)
\end{aligned}$$

We will now prove this. There are two ways to do so, either through recursive substitution or by using the lag operator.

AR(1) to MA(∞) by recursive substitution

We can solve $y_t = \phi y_{t-1} + \varepsilon_t$ with recursive substitution (starting at some infinite time j).

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \varepsilon_t \\
&= \phi\big(\phi y_{t-2} + \varepsilon_{t-1}\big) + \varepsilon_t \\
&= \phi^2(\phi y_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\
&\Downarrow \\
&= \phi^{j+1} y_{t-(j+1)} + \phi^j \varepsilon_{t-j} + \cdots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t
\end{aligned}
$$

The final line explains y as a linear function of an initial value y_{t-j-1} and the historical values of ε_t. If $|\phi| < 1$ and j becomes large, $\phi^{j+1} y_{t-(j+1)} \to 0$. Thus, the AR(1) can be expressed as an MA(∞).

$$
y_t = \phi^j \varepsilon_{t-j} + \cdots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}
$$

AR(1) to MA(∞) with lag operators

Using the lag operator, $y_t = \phi y_{t-1} + \varepsilon_t$ can be written as:

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

A sequence $\{y_t\}_{t=-\infty}^{\infty}$ is said to be bounded if there exists a finite number k, such that $|y_t| < k$ for all t. Provided $|\phi| < 1$ and we restrict ourselves to bounded sequences, the expression can be multiplied by $(1 - \phi L)^{-1}$ on both sides of the equality sign to obtain:

$$y_t = (1 - \phi L)^{-1} \varepsilon_t$$

and solved such that:

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \cdots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

Importantly, we can only perform this inversion (multiplication by $(1 - \phi L)^{-1}$) if $|\phi| < 1$. However, when it holds, we say that the process is invertible.

AR(1) to MA(∞) with lag operators cont'd

How does this work? When $|\phi| < 1$, and $t \to \infty$, the following approximation holds:

$$(1-\phi L)^{-1} = \lim_{j\to\infty}(1+\phi L+(\phi L)^2+\cdots+(\phi L)^j)$$

which is based on the expansion $(1-z)^{-1} = 1+z+z^2+z^3+\cdots$, i.e. the geometric rule. Thus:

$$y_t = (1-\phi L)^{-1}\varepsilon_t$$

can be written as:

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots = \sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j}$$

Mean, variance and autocovariance of an AR(1)

The mean is computed as:

$$\mathrm{E}\left[y_t\right] = \mathrm{E}\left[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots\right] = 0$$

The variance is:

$$\begin{aligned}\gamma(0) = \operatorname{var}\left(y_t\right) &= \mathrm{E}\left[y_t - \mathrm{E}\left[y_t\right]\right]^2 \\ &= \frac{1}{1-\phi^2}\sigma^2\end{aligned}$$

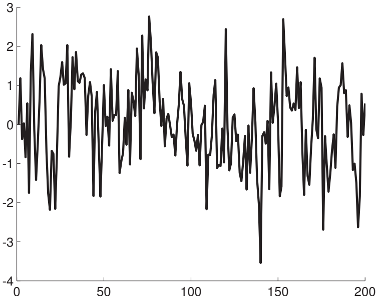
The first order covariance is:

$$\begin{aligned}\gamma(1) &= \mathrm{E}\left[\left(y_t - \mathrm{E}\left[y_t\right]\right)\left(y_{t-1} - \mathrm{E}\left[y_{t-1}\right]\right)\right] \\ &= \phi\frac{1}{1-\phi^2}\sigma^2 = \phi\operatorname{var}\left(y_t\right)\end{aligned}$$

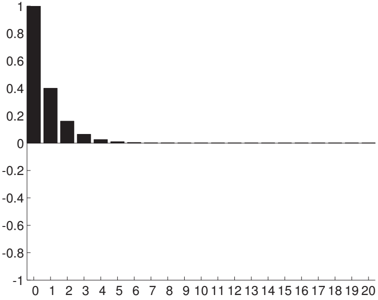
while for $j > 1$ we have:

$$\gamma(j) = \mathrm{E}\left[\left(y_t - \mathrm{E}\left[y_t\right]\right)\left(y_{t-j} - \mathrm{E}[y_{t-j}]\right)\right] = \phi^j\operatorname{var}\left(y_t\right)$$

Figure: AR(1) and the autocorrelation function (ACF)

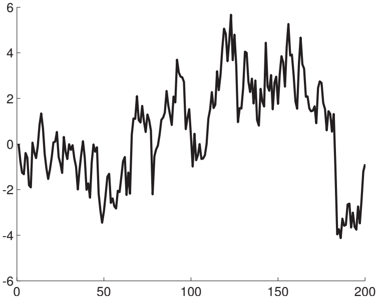


AR $\phi = 0.4$

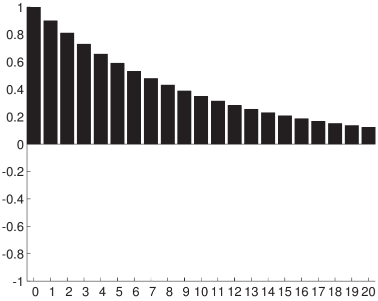


Autocorrelation function (ACF)

Figure: AR(1) and the autocorrelation function (ACF)

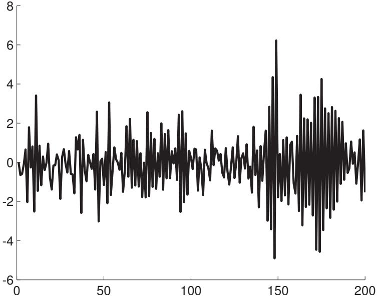


AR $\phi = 0.9$

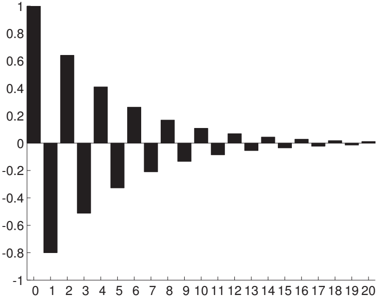


Autocorrelation function (ACF)

Figure: AR(1) and the autocorrelation function (ACF)



AR $\phi = -0.8$



Autocorrelation function (ACF)

Adding a constant

Let us now add a constant to the AR(1) model, so that we can write:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \tag{20}$$

We can easily prove that the results will still be the same, except for the mean, that is no longer zero. To see this, define $\nu_t = \mu + \varepsilon_t$, then:

$$\begin{aligned} y_t &= \phi y_{t-1} + \nu_t \\ y_t &= (1-\phi L)^{-1} \nu_t \\ y_t &= \left(\frac{1}{1-\phi}\right)\mu + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots \end{aligned} \tag{21}$$

Accordingly, the unconditional first moment is:

$$\mathrm{E}\left[y_t\right] = \left(\frac{1}{1-\phi}\right)\mu \tag{22}$$

which is no longer zero.

2. Higher order AR processes

The results for an MA(1) process generalize to an infinite MA. For higher order AR processes, things become a bit more complicated.

Assume an AR(2) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{23}$$

Whether this process is stationary no longer hinges on ϕ_1 alone, as we now also have to consider ϕ_2. As it turns out, there are different ways of doing this.

Higher order AR processes cont.

First, rewrite and define define the (2×1) vector Z_t:

$$Z_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} \tag{24}$$

the (2×1) vector v_t by:

$$v_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{25}$$

and the (2×2) matrix Γ by:

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \tag{26}$$

Higher order AR processes cont.

Then the following first-order vector difference equation can be written as:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{27}$$

or

$$Z_t = \Gamma Z_{t-1} + v_t \tag{28}$$

Γ is called the companion form matrix. Any higher order AR process can be represented as a first-order AR process using this formulation and generalizations of it. We can check stationarity by computing the eigenvalues of matrix Γ. That is, the eigenvalues of Γ are those numbers λ for which $|\Gamma - \lambda I| = 0$. Substituting in for (27), the eigenvalues are the solution to:

$$|\begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}| = 0 \tag{29}$$

$$|\begin{bmatrix} (\phi_1 - \lambda) & \phi_2 \\ 1 & -\lambda \end{bmatrix}| = \lambda^2 - \phi_1 \lambda - \phi_2 = 0 \tag{30}$$

Higher order AR processes cont.

These eigenvalues (λ_1 and λ_2) can be found from the formula:

$$\lambda_1, \lambda_2 = \frac{\left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right)}{2} \tag{31}$$

Stationarity requires that the eigenvalues are less than one in absolute value. In the AR(2) case, one can show that this will be the case if:

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ -\phi_1 + \phi_2 &< 1 \\ \phi_2 &> -1 \end{aligned} \tag{32}$$

see, e.g., Harvey (1993) for proofs.

Autoregressive moving average (ARMA) process

By combining the MA and AR processes, we get and ARMA process. In the simplest case, we can specify an ARMA(1,1) process, which would equal:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

Using the lag polynomial, a more general formulation of an ARMA model is:

$$\phi(L) y_t = \theta(L) \varepsilon_t$$

where

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i \quad \text{and} \quad \theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i$$

The number of lags, p and q, can differ. For instance, an ARMA(2,1) combines an AR(2) with an MA(1):

$$(1 - \phi_1 L - \phi_2 L^2) y_t = (1 + \theta_1 L) \varepsilon_t$$
$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Whether an ARMA process is stationary depends solely on its autoregressive part.

Estimation

An AR(p) model can be estimated using the OLS estimator. An estimate of the population parameters ϕ_p for $p = 1, \dots, P$ and σ^2 can be found by using the OLS quantities $\hat{\phi}_p$, and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t$, where $\hat{\varepsilon}_t = y_t - \sum_{p=1}^{P} \hat{\phi}_p y_{t-p}$. But note: This is a valid procedure only when the process is stationary. Recall the OLS assumption to produce unbiased estimates:

$$E(\varepsilon_t | y_1, \dots, y_T) = 0 \text{ for } t = 1, \dots, T$$

In an autoregressive model like

$$y_t = \alpha + \phi_1 y_{t-1} + \varepsilon_t$$

this does not hold. Clearly, if $\phi_1 \neq 0$, $E(\varepsilon_t | y_{t+1}) \neq 0$, and the relationship breaks down. Then the OLS estimator of ϕ_1 will be biased.

Using large sample theory it can be shown that valid OLS estimation and inference can be conducted when modeling stationary time series.

Lag selection

Alternative methods for determining the lag length are often based on minimizing an information criterion. Two popular information criterion functions are the Bayes (BIC) and the Akaike (AIC) information criterion:

$$BIC(p) = ln(\frac{SSR(p)}{T}) + (p+1)\frac{ln(T)}{T}$$

and

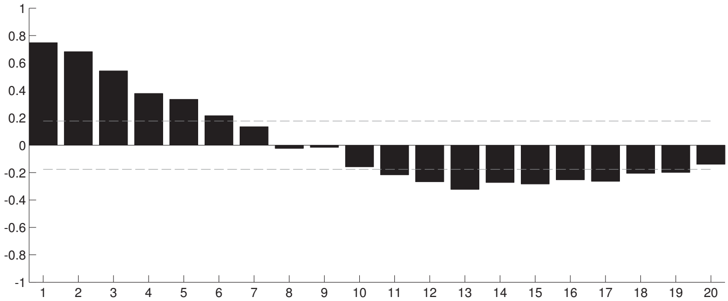
$$AIC(p) = ln(\frac{SSR(p)}{T}) + (p+1)\frac{2}{T}$$

where $SSR(p) = \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t$. The first term will typically decrease as p increases, while the second term increases as the model grows. Difference is in the last term: BIC is more conservative, 'penalizes' the size of the model more.

In practice, neither the BIC nor the AIC (or a regular t-statistic), will likely provide us with one answer. Therefore, depending on the application and the question being asked, the criterion we use will differ.

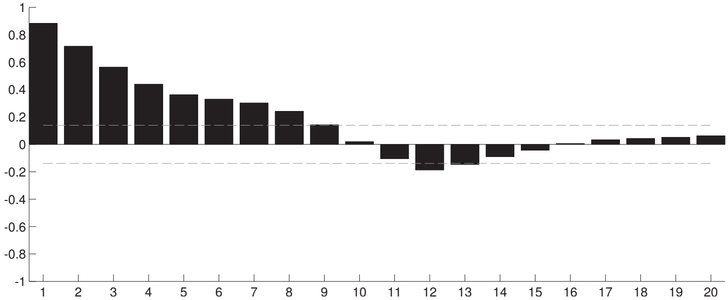
4. An example: GDP and inflation

Figure: Sample autocorrelation function



GDP growth (YoY)

Figure: Sample autocorrelation function



Inflation (YoY)

Table: Estimation results: AR with GDP growth (YoY)

Lag selection test

Criteria	**Lags and scores**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
AIC	0.62	0.56	0.56	0.53	0.52	0.54	0.54	0.46
BIC	0.66	0.62	0.65	0.64	0.66	0.69	0.72	0.66

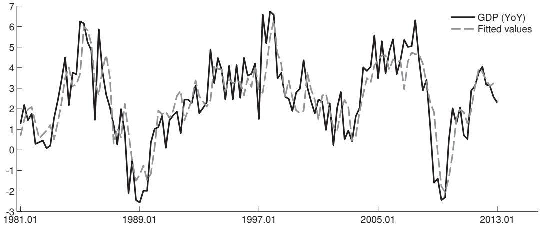
AR(8) estimation results

	Parameters								
	α	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
Estimate	0.76	0.55	0.39	0.01	-0.37	0.21	0.13	0.07	-0.30
P-value	0.00	0.00	0.00	0.90	0.00	0.05	0.23	0.47	0.00

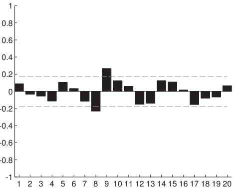
Number of observations:	129
Sample:	1981:Q1 - 2013:Q1
R^2**:**	0.67

Note: See text for details.

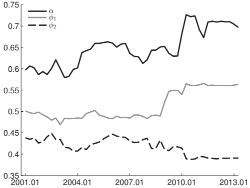
Figure: Estimation results: GDP growth (YoY)



Actual and fitted



Sample autocorrelation function residuals



Parameter (in)stability

Table: Estimation results: AR with inflation (YoY)

Lag selection test

Criteria	**Lags and scores**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
AIC	-1.29	-1.39	-1.38	-1.37	-1.37	-1.36	-1.36	-1.38
BIC	-1.25	-1.34	-1.31	-1.29	-1.27	-1.25	-1.23	-1.23

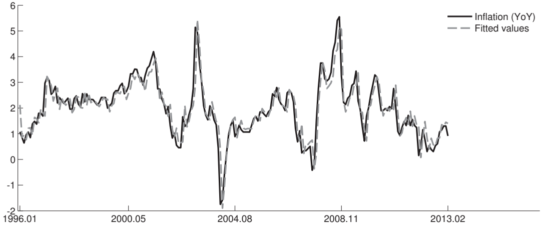
AR(2) estimation results

	Parameters								
	α	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
Estimate	0.29	1.17	-0.32						
P-value	0.00	0.00	0.00						

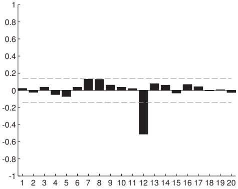
Number of observations:	206
Sample:	1996:M1 - 2013:M2
R^2**:**	0.81

Note: See text for details.

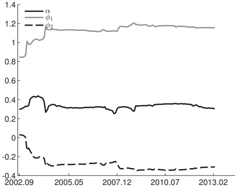
Figure: Estimation results: Inflation (YoY)



Actual and fitted



Sample autocorrelation function residuals



Parameter (in)stability

Multipliers and impulse responses

In time series, we often want to study causes (i.e., an unexpected movement in a given variable or a shock) and the effect (the response to the shock). Assume now an autoregressive AR(1) process, that can be written as an infinite moving average equation:

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \tag{33}$$

Thus, y_t can be described entirely by its past and present errors, or shocks. If we assume the dynamic simulation started at time j, (taking $y_{t-(j+1)}$ as given), the effect of a change in the initial shock (assuming the rest remains the same) on y_t is:

$$\frac{\partial y_t}{\partial \varepsilon_{t-j}} = \phi^j \tag{34}$$

which we call the dynamic multiplier. It depends only on j, the length of time separating the disturbance to the input (ε_{t-j}) and the observed value of output (y_t). It does not depend on time.

Multipliers and impulse responses cont.

The cumulative effect of this temporary shock (we will show later) is

$$\sum_{j=0}^{\infty} \frac{\partial y_t}{\partial \varepsilon_{t-j}} = 1 + \phi + \phi^2 + \cdots + \phi^j = \frac{1}{(1-\phi)} \tag{35}$$

When $|\phi| < 1$ the process decays geometrically towards zero, (positive coefficients – smooth decay; negative coefficients – decay with alternating signs). We say that a system described in this way is stable.

The dynamic multipliers can easily be moved forward in time such that

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = \phi^j$$

Stated like this, the dynamic multipliers for $j = 1, \ldots, J$ are often referred to as the impulse response function. Impulse responses are important tool that enables us to study the effects of structural shocks over time.

Summary

1) Two important time series models are the moving average (MA) and the autoregressive (AR) model.

2) The MA model is stationary regardless of its parameters. The AR model is stationary if the absolute value of the eigenvalues associated with the companion form matrix are all less than one.

3) Under the maintained assumption of stationarity, an AR model can be represented as an infinite moving average model.

4) Estimation of stationary AR models can be conducted using OLS.

5) Different information criteria, like the AIC or BIC, can be used to choose the appropriate lag length.