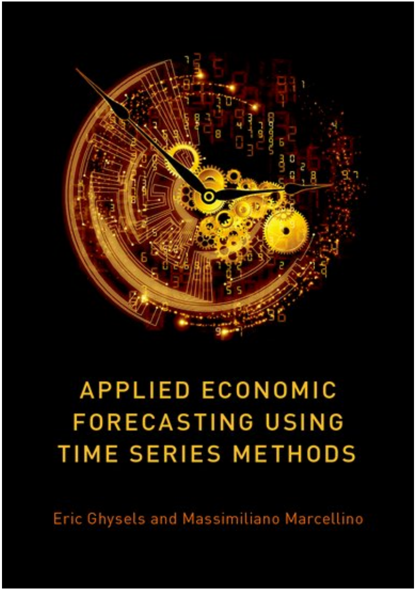

APPLIED ECONOMIC
FORECASTING USING
TIME SERIES METHODS
Eric Ghysels and Massimiliano Marcellino

Applied Economic Forecasting using Time Series Methods

Eric Ghysels and Massimiliano Marcellino

Companion Slides - Chapter 5 Univariate Time Series Models

Overview

- Box and Jenkins popularized the use of univariate time series models for forecasting.
- The key idea is to exploit the past behavior of a time series to forecast its future development, which requires the future to be rather similar to the past.
- In this chapter we will consider the specification, estimation, diagnostic checking, and use for forecasting of ARMA(p,q) models.

The Baseline Linear Regression Model - Outline

Stationarity

- A time series process is **strictly stationary** if

$$\mathcal{F}\left\{y_t, \ldots, y_{t+T}\right\} = \mathcal{F}\left\{y_{t+k}, \ldots, y_{t+T+k}\right\}, \qquad \forall\ t, T, k. \tag{1}$$

  where $\mathcal{F}(\cdot)$ indicates the joint density for a segment of length T of the process y.

- A time series process is **weakly stationary** if

$$\begin{aligned} E\left(y_t\right) &= E\left(y_{t+k}\right), \qquad \forall\ t, k, \\ Var\left(y_t\right) &= Var\left(y_{t+k}\right), \qquad \forall\ t, k, \\ Cov\left(y_t, y_{t-m}\right) &= Cov\left(y_{t+k}, y_{t-m+k}\right), \qquad \forall\ t, m, k. \end{aligned} \tag{2}$$

Stationarity

- A strictly stationary process is also a weakly stationary process provided $\mathcal{F}(\cdot)$ has finite first and second moments.
- A weakly stationary process is not in general strictly stationary.

Representation

- A weakly stationary process can be represented as:

$$\begin{aligned} y_t &= \varepsilon_t + c_1\varepsilon_{t-1} + c_2\varepsilon_{t-2} + \ldots \\ &= \sum_{i=0}^{\infty} c_i \varepsilon_{t-i} = \sum_{i=0}^{\infty} c_i L^i \varepsilon_t \\ &= c(L)\,\varepsilon_t, \end{aligned} \tag{3}$$

  - L is the lag operator: $L\varepsilon_t = \varepsilon_{t-1}$ and $L^i\varepsilon_t = \varepsilon_{t-i}$
  - $c_0 = 1$
  - the error process ε_t is uncorrelated across time and has a constant variance, hence: $\varepsilon_t \sim$ WN$(0, \sigma^2)$, meaning white noise with mean zero and variance σ^2.
- The infinite moving average representation in (3) is known as the Wold decomposition.

Normalization

- For this chapter we assume that the overall mean of the process y_t is zero.
- This my not be true of all time series. However, if we denote the unconditional mean by μ, all the analysis in this chapter goes through with y_t replaced by $(y_t - \mu)$ whenever we have a non-zero mean process.

- A problem with the model in (3) is that it has an infinite number of parameters. However, in general, we can approximate $c(L)$ via a ratio of two finite polynomials, namely:

$$c(L) = \frac{\psi(L)}{\phi(L)}$$

  - $\psi(L)$ = 1 - $\psi_1 L$ - . . . - $\psi_q L^q$
  - $\phi(L)$ = 1 - $\phi_1 L$ - . . . - $\phi_p L^p$

Because of weak stationarity $\phi(L)$ is invertible, i.e., $\phi(z) = 0 = \sum_{j=0}^{p} \phi_j z^j$ has all the roots outside the unit circle. Thus we can rewrite (3) as

$$\phi(L)\, y_t = \psi(L)\, \varepsilon_t,$$

or equivalently as

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \psi_1 \varepsilon_{t-1} - \ldots - \psi_q \varepsilon_{t-q}.$$

- This is a **moving average autoregressive process** of order p and q, henceforth called **ARMA(p,q).**

Autocorrelation

Two useful tools to study ARMA processes are the autocorrelation (AC) and partial autocorrelation (PAC) functions.

- The autocovariance function reports the covariance of y_t with its own lags:

$$\begin{aligned} Cov\left(y_t, y_{t-1}\right) &= \gamma\left(1\right), \\ Cov\left(y_t, y_{t-2}\right) &= \gamma\left(2\right), \\ &\vdots \\ Cov\left(y_t, y_{t-k}\right) &= \gamma\left(k\right). \end{aligned} \tag{4}$$

- Then, the AC is defined as:

$$AC(k) = \frac{Cov\left(y_t, y_{t-k}\right)}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-k})}} = \frac{\gamma\left(k\right)}{\gamma\left(0\right)}.$$

Partial Autocorrelation

The k^{th} value of the PAC measures the correlation between y_t and y_{t-k}, conditional on $y_{t-1}, \ldots, y_{t-k+1}$.

- The elements of the PAC can be considered as specific coefficients in regression equations. In particular, they are

$$\begin{array}{lcl}
PAC(1) & : & \text{coefficient of } y_{t-1} \text{ in the regression of } y_t \text{ on } y_{t-1}, \\
PAC(2) & : & \text{coefficient of } y_{t-2} \text{ in the regression of } y_t \text{ on } y_{t-1}, y_{t-2}, \\
\vdots & & \qquad\vdots \qquad\qquad\qquad \vdots \\
PAC(k) & : & \text{coefficient of } y_{t-k} \text{ in the regression of } y_t \text{ on } y_{t-1}, \ldots, y_{t-k}.
\end{array}$$

- Note that inserting a deterministic component (like a constant) in the model changes the expected value of y_t, while both the AC and PAC remain the same.

Representation

Before studying the characteristics of the class of $ARMA$ processes, we analyze the pure AR and MA processes seperately.

Autoregressive processes

We assume that $c\,(L)$ in (3) is invertible, i.e., $c(z) = 0 = \sum_{j=0}^{\infty} c_j z^j$ has all the roots outside the unit circle. Then we can rewrite (3) as an AR(∞):

$$y_t = \sum_{j=1}^{\infty} \phi_j y_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim \text{WN}(0, \sigma^2). \tag{5}$$

- Weak stationarity implies that the effect of y_{t-j} onto y_t fades as j becomes large.
- In practice we can reasonably approximate (5) with an AR(p):

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t. \tag{6}$$

- If we view the $AR(p)$ process in (6) as an approximation of (3), then it is weakly stationary by definition.
- If instead (6) is the data generating process (DGP), then for y_t to be weakly stationary the roots of $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p = 0$ are all larger than one in absolute value.
- Weak stationarity also guarantees that $\phi(L)$ can be inverted, namely, y_t can be represented as

$$y_t = \frac{1}{\phi(L)} \varepsilon_t, \tag{7}$$

  which is an MA(∞) representation.
- Adding a deterministic component in (6) does not create additional complications.

AR(1) Example

As an example, consider the following $AR(1)$ process:

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ \phi(L) &= 1 - \phi_1 L. \end{aligned} \tag{8}$$

- Weak stationarity requires $|\phi_1| < 1$, since

$$\phi(z) = 1 - \phi_1 z = 0 \rightarrow z = \frac{1}{\phi_1}.$$

AR(1) Example

The $MA\,(\infty)$ of (8) representation is

$$\begin{aligned} y_t &= \varepsilon_t + \phi_1 y_{t-1} \\ &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 y_{t-2} \\ &= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 y_{t-3} \\ &\quad \vdots \\ &= \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} = \frac{1}{\phi(L)} \varepsilon_t. \end{aligned} \tag{9}$$

We can use the MA(∞) representation to compute the mean and variance of of y_t.

$$
\begin{aligned}
E\left(y_t\right) &= 0 \\
Var\left(y_t\right) &= Var\left(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} Var\left(\phi_1^i \varepsilon_{t-i}\right) \\
&= \sum_{i=0}^{\infty} \phi_1^{2i} Var\left(\varepsilon_{t-i}\right) = \frac{\sigma^2}{1-\phi_1^2}
\end{aligned}
$$

AR(1) Example

- The autocovariance function appearing in (5) for the AR(1) is as follows:

$$
\begin{aligned}
\gamma(1) &= Cov\left(\phi_1 y_{t-1} + \varepsilon_t, y_{t-1}\right) = \phi_1 Var\left(y_t\right) = \phi_1 \gamma(0), \\
\gamma(2) &= Cov\left(\phi_1 y_{t-1} + \varepsilon_t, y_{t-2}\right) = \phi_1 \gamma(1) = \phi_1^2 \gamma(0), \\
&\vdots \\
\gamma(k) &= Cov\left(\phi_1 y_{t-1} + \varepsilon_t, y_{t-k}\right) = \phi_1 \gamma(k-1) = \phi_1^k \gamma(0).
\end{aligned}
$$

- Therefore, for $j = 1, 2, \ldots,$ the AC is defined as:

$$
AC(j) = \frac{Cov\left(y_t, y_{t-j}\right)}{\sigma_y \sigma_{y-j}} = \frac{\gamma(j)}{\gamma(0)} = \phi_1^j.
$$

AR(1) Example

- Finally, for the elements of the PAC we have:

$$\begin{aligned} PAC(1) &= \phi_1, \\ PAC(j) &= 0, \quad j > 1. \end{aligned}$$

AR(2) Example

Let us consider now the case of an $AR(2)$ process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t. \tag{10}$$

- To derive the weak stationarity conditions let us consider the solutions of

  $$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0,$$

  which are

  $$z_{1,2} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

AR(2) Example

- Therefore, to have $|z_1| > 1$ and $|z_2| > 1$, we need:

$$\begin{aligned} \phi_1 + \phi_2 &< 1, \\ \phi_2 - \phi_1 &< 1, \\ -\phi_2 &< 1. \end{aligned}$$

AR(2) Example

- The autocovariance function for an AR(2) process is given by:

$$
\begin{aligned}
\gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_\varepsilon^2, \\
\gamma(1) &= \phi_1 \gamma(0) + \phi_2 \gamma(1), \\
\gamma(2) &= \phi_1 \gamma(1) + \phi_2 \gamma(0), \\
&\vdots \\
\gamma(k) &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2).
\end{aligned}
$$

- The general element j of the autocovariance function can be obtained as follows. First, one multiplies both sides of (10) with y_{t-j}. Second, one takes expectations on both sides and uses the fact that y_{t-j} and ε_t are uncorrelated and that $\gamma(i) = \gamma(-i)$.

AR(2) Example

- The γ's can be solved for in terms of the AR parameters using a system of equation for the first three lags of the autocovariance function:

$$
\begin{aligned}
\gamma(0) &= \frac{(1-\phi_1)\,\sigma_\varepsilon^2}{(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]}, \\
\gamma(1) &= \frac{\phi_1\gamma(0)}{1-\phi_2}, \\
\gamma(2) &= \phi_1\gamma(1)+\phi_2\gamma(0),
\end{aligned}
$$

AR(2) Example

- Thus, the autocorrelation function for an AR(2) process is:

$$
\begin{aligned}
AC(1) &= \frac{\phi_1}{(1-\phi_2)} \\
AC(2) &= \phi_2 + \frac{\phi_1^2}{(1-\phi_2)}, \\
&\vdots \\
AC(k) &= \phi_1 AC(k-1) + \phi_2 AC(k-2).
\end{aligned}
$$

- These are known as the Yule-Walker equations, and they can be used to obtain estimators of ϕ_1 and ϕ_2.
- If we substitute $AC(1)$ and $AC(2)$ with their estimated counterparts, the first two equations can be solved for $\hat{\phi}_1$ and $\hat{\phi}_2$.
- There are more efficient estimators than those based on the Yule-Walker equations, but the latter can provide a simple initial estimate.

AR(2) Example

- It can be easily shown that the first two lags of the PAC are different from zero, while $PAC(j) = 0$ for $j > 2$.
- Using methods similar to those seen in the examples considered so far, we can calculate the AC and PAC for any $AR(p)$ process.
- The equations which define the first p lags of the AC function can also be used to obtain the initial (Yule-Walker) estimators of the parameters.

MA Process

The qth-order moving average process, or MA(q), is defined as

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1} - \ldots - \psi_q \varepsilon_{t-q} = \psi(L)\varepsilon_t, \qquad \varepsilon_t \sim WN\left(0, \sigma_\varepsilon^2\right). \tag{11}$$

- It can easily be shown that an MA(q) process is always weakly stationary.
- Its first two moments are

$$\begin{aligned} E(y_t) &= 0, \\ Var(y_t) &= \left(1 + \psi_1^2 + \ldots + \psi_q^2\right)\sigma_\varepsilon^2 = \gamma(0), \end{aligned}$$

- The autocovariance function is as follows:

$$\begin{array}{rcl} \gamma(k) & = & Cov\left(y_t, y_{t-k}\right) \\ & = & Cov\left(\varepsilon_t - \psi_1\varepsilon_{t-1} - \ldots - \psi_q\varepsilon_{t-q}, \varepsilon_{t-k} - \psi_1\varepsilon_{t-k-1} - \ldots - \psi_q\varepsilon_{t-k-q}\right) \\ & = & \left\{ \begin{array}{cc} \left(-\psi_k + \psi_{k+1}\psi_1 + \ldots + \psi_q\psi_{q-k}\right)\sigma_\varepsilon^2 & k = 1, \ldots, q \\ 0 & k > q \end{array} \right. . \end{array}$$

- Dividing $\gamma(k)$ by $\gamma(0)$ yields the autocorrelation function.

MA Process

Another relevant property for an MA process is invertibility, that is, the possibility to represent an MA process as an AR(∞). This requires that all the roots of $\psi(z) = 0$ are larger than one in absolute value.

- When the MA process is invertible, we can write:

$$\frac{1}{\psi(L)} y_t = \varepsilon_t.$$

- The AR(∞) representation is useful to derive the PAC for an MA process, which will coincide with that of the AR(∞) process.
- The PAC elements decay (possibly non-monotonically) towards zero, but are always different from zero except in the limit.

MA(1) Example

As an example, let us consider the MA(1) process

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1}. \tag{12}$$

- Its first two moments are

$$\begin{aligned} E\left(y_t\right) &= 0, \\ Var\left(y_t\right) &= \left(1+\psi_1^2\right)\sigma_\epsilon^2, \end{aligned}$$

- The autocovariance function is

$$\begin{aligned} \gamma(1) &= -\psi_1 \sigma_\epsilon^2 \\ \gamma(k) &= 0, \quad k > 1, \end{aligned}$$

MA(1) Example

- For the AC we have $AC(1) = -\psi_1/(1+\psi_1^2)$ and $AC(k) = 0$, $k > 1$.
- The condition for invertibility is $|\psi_1| < 1$, with associated AR(∞) representation

$$\left(1 + \psi_1 L + \psi_1^2 L^2 + \psi_1^3 L^3 + \ldots\right) y_t = \varepsilon_t.$$

- We can see that the PAC of the $MA(1)$ process declines exponentially.

Remarks on Pure AR and MA Process

Note the different shapes of the AC and PAC for AR and MA processes.

- For an $AR(p)$ process the AC decays but is always different from zero, except in the limit.
- For a $MA(q)$ process the AC is only different from zero up to q lags.
- The PAC of an $AR(p)$ process is only different from zero up to p lags
- The PAC of an $MA(q)$ process decays but is always different from zero, except in the limit.

These considerations suggest that the estimated AC and PAC could be used to determine whether the underlying process is of the AR or MA type, and what is the order of the lags. We will revisit this issue latter.

ARMA Process

An $ARMA(p,q)$ process is defined as

$$\phi(L)\, y_t = \psi(L)\, \varepsilon_t. \tag{13}$$

- The weak stationarity condition is, as for the AR processes,

$$\phi(z) = 0 \rightarrow |z_i| > 1, \qquad i = 1, \ldots, p,$$

  where z_i are the roots of the AR polynomial.
- Likewise, the invertibility condition is, as for the MA processes,

$$\psi(z) = 0 \rightarrow |z_i| > 1, \qquad i = 1, \ldots, q.$$

  where z_i are the roots of the MA polynomial.

ARMA Process

- For a stationary $ARMA$, we can write the infinite MA representation as:
$$y_t = \phi^{-1}(L)\,\psi(L)\,\varepsilon_t = c(L)\,\varepsilon_t$$
- Therefore, the expected value and variance are:
$$E(y_t) = 0 \quad (14)$$
$$Var(y_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} c_i^2. \quad (15)$$
- The AC and the PAC are like those of an $MA(\infty)$ or $AR(\infty)$, so both decline exponentially.

ARMA (1,1) Example

As an example, let us derive the autocovariance structure for an ARMA(1,1) process:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t - \psi_1 \varepsilon_{t-1}. \tag{16}$$

- We can write γ_0 and γ_1 as

$$\begin{aligned} \gamma_0 &= Var\left(y_t\right) = E\left(\phi_1 y_{t-1} + \varepsilon_t - \psi_1 \varepsilon_{t-1}\right)^2 = \\ &= \phi_1^2 \gamma_0 + \sigma_\varepsilon^2 + \psi_1^2 \sigma_\varepsilon^2 - 2\phi_1 \psi_1 E\left(y_{t-1}\varepsilon_{t-1}\right) \\ &= \phi_1^2 \gamma_0 + \sigma_\varepsilon^2 + \psi_1^2 \sigma_\varepsilon^2 - 2\phi_1 \psi_1 \sigma_\varepsilon^2, \end{aligned}$$

and

$$\begin{aligned} \gamma_1 &= Cov\left(y_t, y_{t-1}\right) = E\left(y_{t-1}\left(\phi_1 y_{t-1} + \varepsilon_t - \psi_1 \varepsilon_{t-1}\right)\right) \\ &= \phi_1 \gamma_0 - \psi_1 \sigma_\varepsilon^2 = \frac{\left(1 - \phi_1 \psi_1\right)\left(\phi_1 - \psi_1\right)}{1 - \phi_1^2} \sigma_\varepsilon^2. \end{aligned}$$

ARMA (1,1) Example

- We can write γ_2 as:

$$\gamma_2 = Cov\left(y_t, y_{t-2}\right) = E\left(y_{t-2}\left(\phi_1 y_{t-1} + \varepsilon_t - \psi_1 \varepsilon_{t-1}\right)\right) = \phi_1 \gamma_1.$$

- In general γ_k can be written as:

$$\gamma_k = Cov\left(y_t, y_{t-k}\right) = \phi_1 \gamma_{k-1},$$

- The the kth autocovariance function is

$$AC(k) = \gamma_k / \gamma_0.$$

- Therefore, for $k > 2$ the shape of the AC is similar to that of an $AR(1)$ process.

ARMA (1,1) Example

- It can be easily shown that the MA component plays a similar role for the PAC, in the sense that for $k > 2$ the shape of the PAC is similar to that of an $MA(1)$ process.

ARMA (1,1) Example

Finally, to obtain a first estimate of the $ARMA(1,1)$ parameters ϕ_1, ψ_1, σ_ε^2 we can use the Yule-Walker approach.

- Form a system with the equations for γ_0, γ_1, γ_2.
- Using sample counterparts $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\gamma}_2$, one obtains parameter estimates by solving the system for $\hat{\phi}_1, \hat{\psi}_1, \hat{\sigma}_\varepsilon^2$.

Integrated Process

An **integrated process** y_t is a non stationary process such that $(1-L)^d y_t$ is stationary, where d is the order of integration. The process is typically labeled $I(d)$.

- These are also called unit root processes, since integration is associated with roots of the AR polynomial $\phi(L)$ exactly equal to one, namely, one or more values for z in $\phi(z) = 0$ are equal to one.
- In the weakly stationary case, all the roots must be larger than one in absolute value.

Random Walk

The most common integrated process is the **Random Walk** (RW):

$$y_t = y_{t-1} + \varepsilon_t, \tag{17}$$

for which $d = 1$, since

$$(1 - L)y_t = \Delta y_t = \varepsilon_t.$$

- A RW can also be written as

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots, \tag{18}$$

  so that the effects of a shock do not decay over time, contrary to the case of a weakly stationary process (compare (18) with (9)).

Random Walk

- From (18) we see that
$$E(y_t) = 0.$$
- The variance is not properly defined:
$$Var(y_t) = Var(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots) \to \infty.$$
- The AC is also not properly defined, but the persistence of the effects of the shocks is such that if we computed empirically the AC, its elements would not decay as in the weakly stationary case but would remain close to one at all lags.
- For the PAC, we can easily see that $PAC(1) = 1$, while $PAC(j) = 0$ for $j > 1$.

From (18) we see that

$$E(y_t) = 0,$$

while the variance is not properly defined:

$$Var(y_t) = Var(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots) \to \infty.$$

The AC is also not properly defined, but the persistence of the effects of the shocks is such that if we computed empirically the AC, its elements would not decay as in the weakly stationary case but would remain close to one at all lags. For the PAC, we can easily see that $PAC(1) = 1$, while $PAC(j) = 0$ for $j > 1$.

Random Walk with a Drift

Inserting a deterministic component into an integrated process can change its features substantially. As an example, let us consider the RW with drift:

$$y_t = \mu + y_{t-1} + \varepsilon_t. \tag{19}$$

Random Walk with a Drift

- Repeated substitution yields:

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

so that

$$\begin{aligned} E\left(y_t\right) &= E(\mu + \varepsilon_t + \mu + \varepsilon_{t-1} + \mu + \varepsilon_{t-2} + \ldots) &\rightarrow& \quad \infty, \\ Var\left(y_t\right) &= Var(\mu + \varepsilon_t + \mu + \varepsilon_{t-1} + \mu + \varepsilon_{t-2} + \ldots) &\rightarrow& \quad \infty, \end{aligned}$$

and the AC and PAC are as in the RW case.

- Also in this case first differencing eliminates the non-stationarity, since

$$(1 - L)\, y_t = \Delta y_t = \mu + \varepsilon_t.$$

Final Notes

- Recall that throughout the chapter we assumed for convenience that the mean of y_t is zero. It is clear that from the above discussion that this assumption can be maintained for integrated processes once we subtract μ from $(1 - L)\, y_t$.

ARIMA Process

An $ARIMA(p,d,q)$ process is

$$\phi(L)\,\Delta^d y_t = \psi(L)\,\varepsilon_t. \tag{20}$$

with $\Delta^d \equiv (1-L)^d$, whereas $\phi(L)$ and $\psi(L)$ are polynomials in the lag operator of order, respectively, p and q, while y_t is $I(d)$.

- If we define $x_t = \Delta^d y_t$, x_t is an $ARMA(p,q)$.
- Hence, the only additional complication with respect to the $ARMA$ case is the determination of the order of integration d.

Model Specification

In order to specify an ARIMA model, we need to determine d, p and q, namely, the order of integration and the lag length of the AR and MA components. Three main approaches are available, based respectively on the AC/PAC, testing, and information criteria.

AC/PAC Based Specification

Using the available data on y_t, we can easily estimate the AC and PAC for different values of k. These values are typically graphed, with increasing values of k on the right axis.

- If the estimated values of both AC and PAC decay when k increases, there is evidence that y_t is weakly stationary and we can set $d = 0$.
- If the AC declines very slowly and PAC(1) is close to one, then there is evidence for a unit root. Therefore, we difference y_t once, and repeat the analysis with Δy_t. If necessary, we further difference y_t, otherwise we set $d = 1$ and move to determine p and q.

AC/PAC Based Specification

- Evidence of pure AR(p)
  - PAC presents some peaks and then it is close to zero.
  - AC declines exponentially.
  - p is equal to the number of peaks (coefficients statistically different from zero) in the PAC.
- Evidence of pure MA(q)
  - AC present peaks and then it is close to zero.
  - PAC declines exponentially.
  - q is equal to the number of peaks in the AC.

AC/PAC Based Specification

- Often, we have both AR and MA components (an ARMA process).
  - The AC and PAC can provide an idea on the order of the AR and MA components.
  - Identification of p and q from AC/PAC is more complex in the ARMA case.
  - One possibility is to make a guess on p and q, estimate the ARMA model, control whether the resulting residuals are WN, and if not go back and try with a higher value for p and/or q.

Testing Based Specification

Another diagnostic tool is testing for autocorrelations in the residuals. If we have a correctly specified an $ARMA(p,q)$ model, then we should expect that the estimated errors, $\hat{\varepsilon}_t$, are temporally uncorrelated. Two tests are:

- Ljung-Box Q test (Ljung and Box (1978))
- Box-Pierce test (Box and Pierce (1970))

For both tests, the null hypothesis is that the correlations in the population from which the sample is taken are zero

Testing Based Specification

- The Ljung-Box Q test is defined as:

$$Q_{LB} = T\,(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k}$$

- The Box-Pierce test is defined as:

$$Q_{BP} = T\sum_{k=1}^{h}\hat{\rho}_k^2,$$

  - $\hat{\rho}_k$ is the sample autocorrelation at lag k,.
  - h is the number of lags being tested.
  - For both, under the null Q follows a $\chi^2_{(h)}$.
  - The degrees of freedom should be set to $h-p-q$.

Testing Based Specification

- Many regression-based model selection procedures exist for ARIMA(p,d,0) type models, i.e., models without MA terms, based on formal testing procedures, such as the Wald or LR statistics.
- Testing the lag length of the MA component is more difficult.
  - Its presence prevents the use of OLS estimation.

We can also test for ARCH (autoregressive conditional heteroskedasticity).

- The test involves regressing ε_t^2 onto a constant and h lags ε_{t-1}^2 ..., ε_{t-h}^2.
- Using a straightforward derivation of the LM test leads to the TR^2 test statistic, where the R^2 pertains to the aforementioned regression.
- Under the null hypothesis that there is no ARCH, the test statistic is asymptotically distributed as chi-square distribution with h degrees of freedom.

Specification with Information Criteria

The third option for ARIMA model specification is information criteria (IC), which combine goodness of fit with a penalty function related to the number of model parameters.

- Start with an ARIMA(p_{MAX},d, q_{MAX}) wher it is assumed that $p_{MAX} > p_0$ and $q_{MAX} > q_0$, where p_0 and q_0 are the true orders.
- Compute the IC for all combinations of p, q, and possibly d, selecting the model with the lowest IC.

Specification with Information Criteria

A technical condition related to the penalty function guarantees the procedure provides consistent results asymptotically, in the sense of selecting the true orders p_0 and q_0 with probability approaching one when the sample size T diverges.

- This condition is satisfied by the BIC criterion

$$BIC = \log(\widehat{\sigma}_{\varepsilon}^2) + (p+q)\log(T)/T$$

- but not by the AIC criterion

$$AIC = \log(\widehat{\sigma}_{\varepsilon}^2) + 2(p+q)/T$$

- In finite samples the relative performance of the BIC and AIC is not uniquely defined.
- The BIC generally leads to more parsimonious specifications than the AIC

Concluding Remarks on Specification

- The model specification methods can be applied jointly.
- It may be that several model specifications perform similarly, in which case we can proceed to forecasting with several specifications.
- In general, lower orders of p and q are preferred.
- Often, pure AR approximations are used in forecasting exercises.
- The choice of d is not a major practical issue in practice.
  - If $d_0 = 1$ and we set $d = 0$, then an AR root will be close to one.
  - If $d_0 = 0$ and we set $d = 1$, then an MA root will be close to one, canceling the effected of the imposed unit root.
- Precise specification may be more relevant when considering structural models.

The Baseline Linear Regression Model - Outline

Once the order of the ARIMA(p,d,q) is specified, we need to estimate the parameters in $\phi(L)$ and $\psi(L)$, where

$$\phi(L)\,\Delta^d y_t = \psi(L)\,\varepsilon_t.$$

- If we define $w_t = \Delta^d y_t$, then

$$\phi(L)\,w_t = \psi(L)\,\varepsilon_t$$

  and again

$$\varepsilon_t = \psi(L)^{-1}\,\phi(L)\,w_t.$$

Estimation

Suppose the objective function to be minimized is the usual sum of squared errors:

$$S_t = \sum \varepsilon_t^2.$$

- If there is an MA component, S_t is non-linear in the parameters.
- For example, for an MA(1) process we have

$$y_t - \psi_1 y_{t-1} - \psi_1^2 y_{t-2} - \ldots = \varepsilon_t.$$

  - We cannot find an analytical expression for the parameter estimators. Non-linear least squares (NLS) estimators need to be used.

Typically, the estimator used for ARMA models is not NLS, but MLE.

- Make the additional assumptions that $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ for all t.
- Collect all parameters to be estimated into a single vector θ.

Estimation

An additional complication is dealing with the starting values. For example, for an ARMA(1,1) how do we treat y_0 and ε_0 which are in the specification for y_1? Solution following Box and Jenkins (1976, p.211):

- Compute the likelihood of an ARMA(p,q) conditional on $y_1, \ldots, y_p$, equal to the first p observations and $\varepsilon_1 = \ldots = \varepsilon_{Max(p,q)} = 0$.
- The resulting sample log likelihood for an ARMA(p,q) model is therefore:

$$\begin{aligned} \mathcal{L}(y_{p+1}, \ldots, y_T; \theta) &= -\frac{T-p}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) \\ &\quad - \sum_{t=p+1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \end{aligned} \tag{1}$$

Estimation

- Under mild conditions, the MLE estimator is consistent and, when multiplied by $\sqrt{T}$, it has an asymptotically normal distribution, centered at the true parameter values denoted by θ_0.
- There is no analytical expression for the estimator in the presence of an MA component.
- For convergence of a numerical optimization method, we need a reasonable initial value. One solution is to use the Yule-Walker equations for an initial value.

The Baseline Linear Regression Model - Outline

Consider the following model:

$$\begin{aligned} y_t &= T_t + z_t \\ T_t &= \nu_0 + \nu_1 t \\ z_t &= \rho z_{t-1} + \varepsilon_t \end{aligned} \tag{2}$$

with $\varepsilon_t \sim$ WN$(0, \sigma^2)$, and T_t is a deterministic linear trend.

- If $\rho < 1$ then y_t is I(0) about the deterministic trend T_t, sometimes referred to as *trend-stationary*.
- If $\rho = 1$ and $\nu_1 = 0$, then z_t is a random walk and y_t is I(1) with drift.

Unit Root Tests

Let us start with $\nu_0 = \nu_1 = 0$ and hence T_t is zero, yielding:

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{3}$$

or

$$\Delta y_t = (\rho - 1)\, y_{t-1} + \varepsilon_t. \tag{4}$$

- We are interested in whether there is a unit root, i.e. $\rho = 1$. Hence we are interested in the following hypothesis:

$$\begin{array}{llll} H_0: & \rho = 1 & \rightarrow y_t & \text{is I(1) without drift} \\ H_1: & |\rho| < 1 & \rightarrow y_t & \text{is I(0) with mean zero} \end{array}$$

- Under the unit root null, the usual sample moments of y_t used to compute the $t-$test do not converge to fixed constants.
- Instead, Dickey and Fuller (1979) and Phillips (1987) showed that the sample moments of y_t converge to random functions of Brownian motions on the unit interval.In particular:

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^{T} y_{t-1} &\overset{d}{\to} \sigma \int_0^1 W(\tau)d\tau \\
T^{-2}\sum_{t=1}^{T} y_{t-1}^2 &\overset{d}{\to} \sigma^2 \int_0^1 W(\tau)^2 d\tau \\
T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t &\overset{d}{\to} \sigma^2 \int_0^1 W(\tau)dW(\tau)
\end{aligned}$$

Unit Root Tests

- Using these results Phillips showed that under the unit root null:

$$
\begin{aligned}
T(\hat{\rho}-1) &\quad \xrightarrow{d} \quad \frac{\int_0^1 W(\tau)dW(\tau)}{\int_0^1 W(\tau)^2 d\tau} \\
t - test(\rho = 1) &\quad \xrightarrow{d} \quad \frac{\int_0^1 W(\tau)dW(\tau)}{(\int_0^1 W(\tau)^2 d\tau)^{1/2}}
\end{aligned}
$$

- The convergence rate of $\hat{\rho}$ is not the standard $\sqrt{T}$ but instead T, typically referred to as '*super-consistency*.
- Neither $\hat{\rho}$ nor $t - test(\rho = 1)$ are asymptotically normally distributed.
- The limiting distribution of $t - test(\rho = 1)$ is called the Dickey-Fuller (DF) (Dicky and Fuller (1979)) distribution, which does not have a closed form representation. Consequently,
  - The p-values of the distribution must be computed by numerical approximation or by simulation.
  - Critical values are available in tables or are automatically generated by econometric software packages.

- The $T(\hat{\rho}-1)$ and $t-test(\rho=1)$ statistics are called respectively Dickey-Fuller (DF) normalized bias test and DF t-test.
- The critical values of the DF distribution are generally larger than the standard t distribution ones, such that using the standard critical values will lead to rejecting the null hypothesis of a unit root too often.

Including a deterministic trend to the model yields a different distribution of the test statistics. The two most common trend cases are a **constant only** and a **constant and time trend**

Unit Root Tests with Constant Only

- The test regression for a constant only is

$$\Delta y_t = \nu_0 + (\rho - 1)\, y_{t-1} + \varepsilon_t \tag{5}$$

- The hypotheses to be tested is:

$$\begin{array}{llll} H_0: & \rho = 1 \text{ and } \nu_0 = 0 & \rightarrow y_t & \text{is I(1) without drift} \\ H_1: & |\rho| < 1 \text{ and } \nu_0 \neq 0 & \rightarrow y_t & \text{is I(0) with non-zero mean} \end{array}$$

- Under H_0 the asymptotic distributions of the normalized biased and t test statistics are influenced by the presence but not the coefficient value of the intercept in the DF test regression.

Unit Root Tests with Constant and Time Trend

- The test regression for a constant and time trend is:

$$Deltay_t = \nu_0 + \nu_1 t + (\rho - 1)\, y_{t-1} + \varepsilon_t \tag{6}$$

- The hypotheses to be tested are

$$\begin{array}{llll} H_0: & \rho = 1 \text{ and } \nu_1 = 0 & \rightarrow y_t & \text{is I(1) with drift} \\ H_1: & |\rho| < 1 \text{ and } \nu_1 \neq 0 & \rightarrow y_t & \text{is I(0) with deterministic trend} \end{array}$$

- Under H_0 the asymptotic distributions of the normalized biased and t test statistics are again influenced by the presence but not the parameter values of the intercept and trend slope coefficients in the DF test regression.

Unit Root Tests with $p > 1$

If the AR process is of order p , with $p > 1$, we can rewrite the process as

$$\Delta y_t = \gamma y_{t-1} + \rho_1 \Delta y_{t-1} + \ldots + \rho_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \qquad (7)$$

where $\gamma = \phi(1)$, and the ρ coefficients are related to the original ϕ coefficients.

- If one of the roots of $\phi(z) = 0$ is $z = 1$, it follows that $\phi(1) = \gamma = 0$ and therefore the coefficient of y_{t-1} in (7) will be equal to zero.
- We can apply the same test as before, namely a t-test for $\gamma = 0$, or normalized bias test, both known as Augmented Dickey Fuller (ADF) tests.
- The ADF test has the same distribution as the DF test.

Unit Root Tests with $p > 1$

We can run test regressions:

$$\Delta y_t = D_t \beta + \gamma y_{t-1} + \rho_1 \Delta y_{t-1} + \ldots + \rho_{p-1} \Delta y_{t-p+1} + \varepsilon_t \qquad (8)$$

where D_t is a vector of deterministic terms (constant, trend).

- The specification of the deterministic terms depends on the assumed behavior under the alternative hypothesis of trend stationarity.
- The ADF t-statistic and normalized bias statistic are based on the least squares estimates of the above regression and have the same limiting distribution as with $p = 1$.
- In finite samples, the size and power of the ADF test are affected by the number of lags included in (7), so that the determination of p is relevant.

Unit Root Tests

- Other unit root tests include:
  - Phillips-Perron (PP) unit root tests (Phillips and Perron (1988)).
  - Elliot, Rothenburg, and Stock (1996) proposed a modification of the DF test statistic based on the generalized least squares (GLS) principle, known as the DF-GLS.

- The size and power of unit root tests are also affected by the presence of deterministic breaks in the model parameters.
  - More complex statistical procedures to allow for breaks when testing for unit roots, see e.g.m Perron (1989) and Stock (1994).

Unit Root Tests with More than 1 Unit Root

There are also procedures to test for the presence of more than one unit root, e.g., $d = 2$.

- The simplest approach is first to test whether Δy_t has a unit root (so that $d = 2$) and then, if the hypothesis is rejected, whether y_t has a unit root (so that $d = 1$).
- At each step we can use the DF or ADF tests that we have described.
- One should use proper critical values that control for the sequential applications of the procedure.

For forecasting, it is worth mentioning that if y_t is $I(d)$, so that

$$\Delta^d y_t = w_t, \tag{9}$$

then

$$y_t = \Sigma^d w_t, \tag{10}$$

where, as an example,

$$\Sigma w_t = \sum_{i=-\infty}^{t} w_i, \tag{11}$$

$$\Sigma^2 w_t = \sum_{i=-\infty}^{t} \sum_{i=-\infty}^{j} w_j. \tag{12}$$

Conclusions

- There are many alternative procedures available for testing for unit roots (or for stationarity), notably where the null is trend-stationarity see e.g., Kwiatkowski, Philips, Schmidt, and Shin (1992).
- It is difficult in general to outperform the ADF procedure, which has the advantage of being easy to implement and understand.
- For further details, there are a number of excellent surveys one can read, including Campbell and Perron (1991), Stock (1994), and Phillips and Xiao (1998).

The Baseline Linear Regression Model - Outline

Diagnostic Checking

We noted in Section 3 that the Ljung-Box and Box-Pierce statistics can be used to test whether estimated residuals are white noise.

- More informally, we can verify whether the estimated values of the residual AC and PAC lie within the approximate asymptotic 95% confidence bands (around the null hypothesis of zero), which are $\pm 1.96/\sqrt{T}$.
- We can use also the other tests for no correlation in the residuals that we have seen in the context of the linear regression model, as well as those for homoskedasticity and parameter stability, which all have an asymptotic justification.

Diagnostic Checking

- If the model fails to pass the diagnostic checks, we need to re-specify it, for example by increasing the AR or MA orders.
- If the diagnostic checks are passed but some of the model coefficients are not statistically significant, we could also assess whether the diagnostic checks remain satisfactory for a more parsimonious specification.
- After determining the ARIMA model order, estimating its parameters, and verifying the underlying assumptions are not rejected, we can use it for forecasting purposes.

As we have seen in the case of the linear regression model, the optimal forecast of y_{T+h} in the MSFE sense is

$$\widehat{y}_{T+h} = E\left(y_{T+h}\,|y_T, y_{T-1}, \ldots, y_1\right). \tag{13}$$

- We consider optimal linear forecasts for $ARIMA(p,d,q)$ models, which coincide with $E\left(y_{T+h}\,|y_T, y_{T-1}, \ldots, y_1\right)$ if we assume that $\{\varepsilon_t\}$ is normal.
- We also assume for the moment that the ARIMA parameters are known.
- We first discuss the general case, then present some examples, and finally make a set of additional comments related to ARIMA forecasts.

General Formula

To calculate the optimal linear forecast for an ARIMA(p,d,q), we can proceed as follows. We start by defining $\Delta^d y_t = w_t$, so that w_t is ARMA(p,q):

$$w_T = \phi_1 w_{T-1} + \ldots + \phi_p w_{T-p} + \varepsilon_T - \psi_1 \varepsilon_{T-1} - \ldots - \psi_q \varepsilon_{T-q}.$$

- From

  $$w_{T+1} = \phi_1 w_T + \ldots + \phi_p w_{T-p+1} + \varepsilon_{T+1} - \psi_1 \varepsilon_T - \ldots - \psi_q \varepsilon_{T-q+1},$$

  it follows that

  $$\widehat{w}_{T+1} = E\left(w_{T+1} \left| I_T \right.\right) = \phi_1 w_T + \ldots + \phi_p w_{T-p+1} - \psi_1 \varepsilon_T - \ldots - \psi_q \varepsilon_{T-q+1}.$$

- Similarly,

$$
\begin{aligned}
\widehat{w}_{T+2} &= E\left(w_{T+2}\,|I_T\right) = \phi_1 \widehat{w}_{T+1} + \ldots \\
&\quad + \phi_p w_{T-p+2} - \psi_2 \varepsilon_T - \ldots - \psi_q \varepsilon_{T-q+2} \\
&\quad \ldots \\
\widehat{w}_{T+h} &= E\left(w_{T+h}\,|I_T\right) = \phi_1 \widehat{w}_{T+h-1} + \ldots \\
&\quad + \phi_p \widehat{w}_{T-p+h} - \psi_h \varepsilon_T - \ldots - \psi_q \varepsilon_{T-q+h}
\end{aligned}
\tag{14}
$$

where $\widehat{w}_{T-j} = w_{T-j}$ if $j \leq 0$ and there is no MA component for $h > q$.

General Formula

- We obtain the forecast of y_{T+h} summing appropriately those for w_{T+j}, $j = 1, \ldots, h$. For example, for $d = 1$, we have

$$\widehat{y}_{T+h} = y_T + \widehat{w}_{T+1} + \ldots + \widehat{w}_{T+h}. \tag{15}$$

- To clarify the general derivation of the optimal ARIMA forecasts, we now discuss a few examples.

AR(1) Example

We consider a number of specific examples which are relatively simple to work with beginning with the **AR(1)** process:

$$y_t = \phi y_{t-1} + \varepsilon_t. \tag{16}$$

AR(1) Example

- The formula in (14) simplifies to:

$$
\begin{aligned}
\widehat{y}_{T+1} &= \phi y_T, \\
\widehat{y}_{T+2} &= \phi \widehat{y}_{T+1} = \phi^2 y_T, \\
&\dots \\
\widehat{y}_{T+k} &= \phi^k y_T.
\end{aligned}
$$

Since

$$
\begin{aligned}
y_{T+1} &= \phi y_T + \varepsilon_{T+1}, \\
y_{T+2} &= \phi^2 y_T + \varepsilon_{T+2} + \phi \varepsilon_{T+1}, \\
&\dots \\
y_{T+h} &= \phi^h y_T + \varepsilon_{T+h} + \phi \varepsilon_{T+h-1} + \ldots + \phi^{h-1} \varepsilon_{T+1}.
\end{aligned}
$$

AR(1) Example

- The forecast errors are

$$\begin{aligned} e_{T+1} &= \varepsilon_{T+1}, \\ e_{T+2} &= \varepsilon_{T+2} + \phi\varepsilon_{T+1}, \\ &\ldots \\ e_{T+h} &= \varepsilon_{T+h} + \phi\varepsilon_{T+h-1} + \ldots + \phi^{h-1}\varepsilon_{T+1}, \end{aligned}$$

and their variances are

$$\begin{aligned} Var\left(e_{T+1}\right) &= \sigma_{\varepsilon}^{2}, \\ Var\left(e_{T+2}\right) &= \left(1+\phi^{2}\right)\sigma_{\varepsilon}^{2}, \\ &\ldots \\ Var\left(e_{T+k}\right) &= \left(1+\phi^{2}+\ldots+\phi^{2k-2}\right)\sigma_{\varepsilon}^{2}. \end{aligned}$$

- Moreover, we have

$$\begin{aligned} \lim_{h\to\infty} \widehat{y}_{T+h} &= 0 = E\left(y_t\right) \\ \lim_{h\to\infty} Var\left(\widehat{y}_{T+h}\right) &= \frac{1}{1-\phi^2}\sigma_\varepsilon^2 = Var\left(y_t\right) \end{aligned}$$

MA(1) Example

Now Let us consider the case of an **MA(1)** process,

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1} \tag{17}$$

MA(1) Example

- From the general formula in (14), we have

$$\begin{aligned}
\widehat{y}_{T+1} &= -\psi_1 \varepsilon_T, \\
\widehat{y}_{T+2} &= 0, \\
&\dots \\
\widehat{y}_{T+h} &= 0.
\end{aligned}$$

Since

$$\begin{aligned}
y_{T+1} &= \varepsilon_{T+1} - \psi_1 \varepsilon_T, \\
y_{T+2} &= \varepsilon_{T+2} - \psi_1 \varepsilon_{T+1}, \\
&\dots \\
y_{T+h} &= \varepsilon_{T+h} - \psi_1 \varepsilon_{T+h-1}.
\end{aligned}$$

MA(1) Example

- The forecast errors are

$$\begin{aligned} e_{T+1} &= \varepsilon_{T+1} \\ e_{T+2} &= \varepsilon_{T+2} - \psi_1 \varepsilon_{T+1} \\ &\ldots \\ e_{T+k} &= \varepsilon_{T+k} - \psi_1 \varepsilon_{T+k-1}, \end{aligned}$$

with variances

$$\begin{aligned} Var\left(e_{T+1}\right) &= \sigma_{\varepsilon}^2 \\ Var\left(e_{T+2}\right) &= \left(1 + \psi_1^2\right) \sigma_{\varepsilon}^2 \\ &\ldots \\ Var\left(e_{T+h}\right) &= \left(1 + \psi_1^2\right) \sigma_{\varepsilon}^2 \end{aligned}$$

MA(1) Example

- Again, we have

$$\begin{aligned}\lim_{h\to\infty} \widehat{y}_{T+h} &= 0 = E\left(y_t\right)\\ \lim_{h\to\infty} Var\left(\widehat{e}_{T+h}\right) &= \left(1+\psi_1^2\right)\sigma_\varepsilon^2 = Var\left(y_t\right).\end{aligned}$$

ARMA(p,q) Example

Let us now consider an alternative derivation of the optimal forecast for an ARMA(p,q), and show that it is equivalent to that in (14). We can write the model as

$$a(L)y_t = b(L)\varepsilon_t$$

$$a(L) = 1 - \sum_{j=1}^{p} a_j L^j, \;\; b(L) = \sum_{j=0}^{q} b_j L^j, \quad b_0 = 1$$

ARMA(p,q) Example

- Rewriting as an $MA(\infty)$ representation we have

$$y_t = c(L)\varepsilon_t,$$

with $a(z)c(z) = b(z)$, namely,

$$c_k - \sum_{j=1}^{p} a_j c_{k-j} = b_k, \tag{18}$$

ARMA(p,q) Example

- We have that

$$\hat{y}_{T+h} = \sum_{i=0}^{\infty} c_{i+h}\varepsilon_{T-i}.$$

The above equation implies that

$$\hat{y}_{T+h} - \sum_{j=1}^{p} a_j \hat{y}_{T+h-j} = \sum_{i=0}^{\infty} \left(c_{i+h}\varepsilon_{T-i} - \sum_{j=1}^{p} a_j c_{i+h-j}\varepsilon_{T-i}\right). \tag{19}$$

ARMA(p,q) Example

- If we now use (18) in (19), we have

$$\hat{y}_{T+h} - \sum_{j=1}^{p} a_j \hat{y}_{T+h-j} = \sum_{j=0}^{\infty} b_{j+h} \varepsilon_{T-j}, \tag{20}$$

  with $b_{j+h} \equiv 0, j+h > q$, which is indeed equivalent to what we would obtain with (14).
- With this formula we can generate optimal forecasts for all h, p and q.

Random Walk Example

Now consider the random walk

$$y_t = y_{t-1} + \varepsilon_t, \tag{21}$$

- The optimal forecast is

$$\widehat{y}_{T+h} = y_T,$$

for any h, and

$$e_{T+h} = \varepsilon_{T+1} + \varepsilon_{T+2} + \ldots + \varepsilon_{T+h}..$$

Random Walk Example

- The variance of the forecast error is

$$Var\left(e_{T+h}\right) = h\sigma_{\varepsilon}^{2}.$$

- From these expressions, it follows that

$$\begin{aligned}
\lim_{k\to\infty} \widehat{y}_{T+k} &= y_T, \\
\lim_{k\to\infty} Var\left(\widehat{y}_{T+k}\right) &= \infty.
\end{aligned}$$

- More generally, the presence of a unit root in the AR component implies that the variance of the forecast error grows linearly over time, while in the stationary it converges to the unconditional variance of the variable.

Additional Comments

What follows are a few interesting implications of what we have derived:

- Using the $MA(\infty)$ representation, it can be shown that the forecast error is:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h} = \sum_{j=0}^{h-1} c_j \varepsilon_{T+h-j}. \tag{22}$$

  Therefore, even when using an optimal forecast, the h-steps ahead forecast error is serially correlated. In particular, it is an MA(h-1) process.

Additional Comments

- Moreover

$$
\begin{aligned}
E(e_{T+h}) &= 0, \\
\mathrm{Var}(e_{T+h}) &= \sigma_\varepsilon^2 \sum_{j=0}^{h-1} c_j^2, \\
\lim_{h\to\infty} \mathrm{Var}(e_{T+h}) &= \mathrm{Var}(y_t), \mathrm{Var}(e_{T+h+1}) - \mathrm{Var}(e_{T+h}) = \sigma_\varepsilon^2 c_h^2 \geq 0.
\end{aligned}
$$

Thus the forecast error variance increases monotonically with the forecast horizon. This result is no longer necessarily true if the parameters are estimated.

Additional Comments

- If the error ε is Gaussian, so is the forecast error. In particular

$$\frac{y_{T+h} - \hat{y}_{T+h}}{\sqrt{\text{Var}(e_{T+h})}} \sim N(0,1).$$

- We can use this result to construct $(1-\alpha)\%$ interval forecasts as:

$$\left(\hat{y}_{T+h} - c_{\alpha/2}\sqrt{\text{Var}(e_{T+h})}; \hat{y}_{T+h} + c_{\alpha/2}\sqrt{\text{Var}(e_{T+h})}\right)$$

  where $c_{\alpha/2}$ are critical values from the standard normal distribution.

Additional Comments

- From (22), for $h = 1$ we have

$$e_{T+1} = \varepsilon_{T+1}, \tag{23}$$

which can also be read as

$$\varepsilon_{T+1} = y_{T+1} - \hat{y}_{T+1}, \tag{24}$$

which provides an interpretation of the errors in the MA(∞) representation of a weakly stationary process.

Additional Comments

- Consider $\hat{y}_{T+h}$ and $\hat{y}_{T+h+k}$, i.e., forecasts of y_{T+h} and y_{T+h+k} made in period T. From (22) it can be easily shown that:

$$E(e_{T+h}e_{T+h+k}) = \sigma_{\varepsilon}^2 \sum_{j=0}^{h-1} c_j c_{j+k},$$

  so that the forecast errors for different horizons are in general correlated.

Additional Comments

- From (22) and the fact that ε_t is white noise, and considering the predictor $\hat{y}_{T+h}$ as an estimator (hence random), it follows that

$$Cov(\hat{y}_{T+h}, e_{T+h}) = 0.$$

Therefore,

$$Var(y_{T+h}) = Var(\hat{y}_{T+h}) + Var(e_{T+h})$$

and

$$Var(y_{T+h}) \geq Var(\hat{y}_{T+h}).$$

Hence, the forecast is always less volatile than the actual realized value.

The Baseline Linear Regression Model - Outline

Forecasting, Estimated Parameters

We now move on to the case where parameters are estimated.

- If we use consistent parameter estimators, the formulas we have derived for the optimal forecasts remain valid.
- However, there is an increase in the variance of the forecast error due to the estimation uncertainty.
- To illustrate this we focus on only a few examples.

AR(1) with Drift Example

The first case we assess is that of a stationary AR(1) with drift:

$$y_t = \mu + a y_{t-1} + \varepsilon_t. \tag{25}$$

AR(1) with Drift Example

- If the parameters μ and a have to be estimated, by $\hat{\mu}$ and $\hat{a}$, the forecast error for $h = 1$ is:

$$e_{T+1} = \varepsilon_{T+1} + (\mu - \hat{\mu}) + (a - \hat{a})y_T = \varepsilon_{T+1} + (\Theta - \hat{\Theta})' x_T$$

$$x_T = \begin{pmatrix} 1 \\ y_T \end{pmatrix} \qquad \Theta - \hat{\Theta} = \begin{pmatrix} \mu - \hat{\mu} \\ a - \hat{a} \end{pmatrix}$$

and

$$\mathrm{Var}(e_{T+1}) = \sigma_\varepsilon^2 + x_T' \mathrm{Var}(\hat{\Theta}) x_T, \tag{26}$$

where

$$\begin{aligned} \mathrm{Var}(\hat{\Theta}) = \mathrm{Var}\begin{pmatrix} \hat{\mu} \\ \hat{a} \end{pmatrix} &= \sigma_\varepsilon^2 E \begin{bmatrix} T & \sum_{t=1}^T y_{t-1} \\ \sum_{t=1}^T y_{t-1} & \sum_{t=1}^T y_{t-1}^2 \end{bmatrix}^{-1} \\ &\cong T^{-1} \begin{bmatrix} \sigma_\varepsilon^2 + \mu^2(1+a)(1-a)^{-1} & -\mu(1+a) \\ -\mu(1+a) & (1-a^2) \end{bmatrix}, \end{aligned} \tag{27}$$

AR(1) with Drift Example

- The h-steps ahead prediction is:

$$\hat{y}_{T+h} = \hat{\mu}\frac{(1-\hat{a}^h)}{(1-\hat{a})} + \hat{a}^h y_T.$$

- Therefore,

$$\hat{e}_{T+h} = \sum_{i=0}^{h-1}(\mu a^i - \hat{\mu}\hat{a}^i) + (a^h - \hat{a}^h)y_T + \sum_{i=0}^{h-1} a^i \varepsilon_{T+h-i}$$

$$\begin{aligned}\text{Var}(\hat{e}_{T+h}) \quad = \quad & \sigma_\varepsilon^2 \frac{(1-a^{2h})}{(1-a^2)} + E\left[\sum_{i=0}^{h-1}(\mu a^i - \hat{\mu}\hat{a}^i)\right]^2 \\ & + \text{Var}\left[(a^h - \hat{a}^h)\right] y_T^2 + 2E\left[\sum_{i=0}^{h-1}(\mu a^i - \hat{\mu}\hat{a}^i)(a^h - \hat{a}^h)\right] y_T.\end{aligned}$$

Unit Root Example

Consider a unit root with a drift by setting $a = 1$ in equation (25).

- We have already seen that

$$\hat{y}_{T+h} = \mu h + y_T$$
$$e_{T+h} = \sum_{i=0}^{h-1} \varepsilon_{T+h-i}$$

and $\text{Var}(e_{T+h}) = h\sigma_\varepsilon^2$ increases with the forecast horizon h.

AR(1) with Drift Example

- Using estimated parameters from an AR(1) with drift, without imposing the unit root, the forecast error becomes

$$e_{T+h} = (\mu - \hat{\mu})h + (1 - \hat{a}^h)y_T + \sum_{i=0}^{h-1} \varepsilon_{T+h-i}. \tag{28}$$

- It can be shown that the OLS estimator of a converges at a rate of $T^{3/2}$, rather than $T^{1/2}$ as in the stationary case, so that its variance can be neglected.
- The variance of $\hat{\mu}$ decreases instead with T, so that the key determinant of the forecast uncertainty is the estimation of μ, the "local trend," combined with the cumulated future errors.

The Baseline Linear Regression Model - Outline

Multi-steps (or Direct) Estimation

The idea of **multi-steps (or direct) estimation** is to estimate the parameters that will be used in forecasting by minimizing the same loss function as in the forecast period.

- As an example, let us consider the $AR(1)$:

$$y_t = ay_{t-1} + \varepsilon_t$$

so that

$$y_{T+h} = a^h y_T + \sum_{i=0}^{h-1} a^i \varepsilon_{T+h-i}$$

Standard or Iterated Forecast

- The standard forecast is:

$$\widehat{y}_{T+h} = \widehat{a}^h y_T,$$

where

$$\widehat{a} = \underset{a}{\text{argmin}} \sum_{t=1}^{T} (y_t - a y_{t-1})^2 = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \text{and} \tag{29}$$

$$E(y_{T+h} - \widehat{y}_{T+h}) = (a^h - E(\widehat{a}^h)) y_T. \tag{30}$$

- The forecast $\widehat{y}_{T+h}$ is also called "iterated" as it can be derived by replacing the unknown future values of y with their forecasts for $T+1, \ldots, T+h-1$.

Direct Forecast

The alternative forecast is

$$\widetilde{y}_{T+h} = \widetilde{a}_h y_T$$

where

$$\widetilde{a}_h = \underset{a_h}{\text{argmin}} \sum_{t=1}^{T} (y_t - a_h y_{t-h})^2 = \frac{\sum_{t=h}^{T} y_t y_{t-h}}{\sum_{t=h}^{T} y_{t-h}^2}$$

and

$$E(y_{T+h} - \widetilde{y}_{T+h}) = (a^h - E(\widetilde{a}_h))y_T.$$

- The forecast $\widetilde{y}_{T+h}$ is labeled “direct” since it is derived from a model where the target variable y_{T+h} is directly related to the available information set in period T.

Comparison

How do the forecasts $\widehat{y}_{T+h}$ and $\widetilde{y}_{T+h}$ compare?

- The relative performance in terms of bias and efficiency depends on the bias and efficiency of the alternative estimators of a^h, $\widehat{a}^h$ and $\widetilde{a}_h$.
- In the presence of correct model specification, both estimators of a^h are consistent, but $\widehat{a}^h$ is more efficient than $\widetilde{a}_h$ since it coincides with the maximum likelihood estimator.
- In the presence of model mis-specification the ranking can change.

Example

As an example, assume that the DGP is an MA(1):

$$y_t = \varepsilon_t + \psi \varepsilon_{t-1} \qquad (31)$$
$$\varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$$

- Suppose that the chosen model for y_t is the $AR(1)$:

$$y_t = \rho y_{t-1} + v_t, \qquad (32)$$
$$v_t \sim \text{WN}(0, \sigma_v^2)$$

- For illustration, we will compare standard and direct estimation based forecasts, assuming $h = 2$ and using the MSFE as a comparison criterion.

Example

- Standard estimation yields

$$\hat{\rho} = \sum_{t=1}^{T} y_t y_{t-1} \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1}$$

- And to a first approximation

$$E(\hat{\rho}) \cong \frac{\psi}{(1+\psi^2)} = \rho.$$

Example

- Then

$$\hat{y}_{T+2} = \hat{\rho}^2 y_T, \quad E(\hat{y}_{T+2}) \cong \rho^2 y_T$$

- It can be shown that

$$\begin{aligned}\widehat{MSFE} &= E\left[(y_{T+2} - \hat{\rho}^2 y_T)^2 | y_T\right] \\ &\cong (1+\psi^2)\sigma_\varepsilon^2 + \left(\text{Var}(\hat{\rho}^2) + \rho^4\right) y_T^2\end{aligned}$$

Example

- In the case of direct estimation:
  - $\tilde{\rho}_2 = \sum_{t=2}^{T} y_t y_{t-2} (\sum_{t=2}^{T} y_{t-2}^2)^{-1} \cong 0$
  - $\tilde{y}_{T+2} = \tilde{\rho}_2 y_T \cong 0$
  - $\widetilde{MSFE} = E[(y_{T+2} - \tilde{y}_{T+2})|y_T] \cong (1 + \psi^2)\sigma_\varepsilon^2 + \text{Var}(\tilde{\rho}_2) y_T^2$

Example

- It can be shown that for particular values of the parameters it is possible that
$$\widetilde{MSFE} \leq \widehat{MSFE}.$$
- A necessary condition for this is that the AR(1) model is mis-specified. Otherwise, $\hat{\rho}^2$ is the ML estimator of ρ^2 and the associated forecast cannot be beaten.
- The relative performance of the iterated forecasts improves with the forecast horizon.

The Baseline Linear Regression Model - Outline

Permanent-transitory Decomposition

It is sometimes of interest to decompose a process y_t into two components:

- **Permanent**- captures the long-run, trend-like behavior of y_t.
- **Transitory**- measures short term deviations from the trend.

There is no unique way to achieve a permanent-transitory. We consider the to most common approaches:

- The **Beveridge and Nelson** (1981) (BN) decomposition features a permanent component which behaves as a random walk.
- The **Hodrick and Prescott** (1997) (HP) has potentially more complex dynamics for the permanent component.

Beveridge and Nelson Decomposition

Recall that a weakly stationary process can be written as an MA(∞), and that if $y_t \sim I(d)$, then $\Delta^d y_t$ is weakly stationary.

- Assuming $d = 1$, we have:

$$\Delta y_t = \mu + c(L)\,\varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \left(0, \sigma_\varepsilon^2\right). \tag{33}$$

- Let us consider another polynomial in L, defined as

$$d\left(L\right)=c\left(L\right)-c\left(1\right). \tag{34}$$

- Since $d\left(1\right)=0$, 1 is a root of $d\left(L\right)$, which can therefore been rewritten as

$$d\left(L\right)=\tilde{c}\left(L\right)\left(1-L\right). \tag{35}$$

- Combining equations (34) and (35) we obtain

$$\begin{aligned} (L) &= \tilde{c}\left(L\right)\left(1-L\right)+c\left(1\right) \text{ and} \\ \Delta y_t &= \mu+\tilde{c}\left(L\right)\Delta\varepsilon_t+c\left(1\right)\varepsilon_t. \end{aligned}$$

Beveridge and Nelson Decomposition

- To obtain a representation for y_t, we need to integrate both sides:

$$y_t = \underbrace{\mu t + c(1) \sum_{j=1}^{t} \varepsilon_j}_{\substack{\text{trend} \\ \text{(permanent component)} \\ \text{(PC)}}} + \underbrace{\tilde{c}(L)\, \varepsilon_t}_{\substack{\text{cycle} \\ \text{(transitory component)} \\ \text{(CC)}}} .$$

- It also follows that the permanent component is a random walk with drift:

$$PC_t = PC_{t-1} + \mu + c(1)\, \varepsilon_t. \tag{36}$$

- Moreover, the variance of the trend innovation is $c\,(1)^2\,\sigma_\varepsilon^2$, which is larger (smaller) than the innovation in y_t if $c\,(1)$ is larger (smaller) than one.
- The innovation in the cyclical component is $\tilde{c}\,(0)\,\varepsilon_t$. Since $\tilde{c}\,(L) = \frac{c(L)-c(1)}{1-L}$, then $\tilde{c}\,(0) = c\,(0) - c\,(1) = 1 - c\,(1)$. Therefore, the innovation in the cyclical component is $(1 - c\,(1))\,\varepsilon_t$.

BN Decomposition Example

As an example, let us derive the BN decomposition for an ARIMA(1,1,1) model,

$$\Delta x_t = \phi \Delta x_{t-1} + \varepsilon_t + \psi \varepsilon_{t-1}.$$

- From the MA representation for Δx_t, we have:

$$\begin{aligned} c(L) &= \frac{1+\psi L}{1-\phi L}, \qquad c(1) = \frac{1+\psi}{1-\phi}, \\ \tilde{c}(L) &= \frac{c(L)-c(1)}{1-L} = -\frac{\phi+\psi}{(1-\psi)(1-\psi L)}; \end{aligned}$$

BN Decomposition Example

- It follows that the BN decomposition is

$$y_t = PC + CC = \frac{1+\psi}{1-\phi}\sum_{j=1}^{t}\varepsilon_j - \frac{\phi+\psi}{(1-\psi)(1-\psi L)}\varepsilon_t. \tag{37}$$

The Hodrick-Prescott Filter

The Hodrick-Prescott filter is an alternative way to compute the permanent component. The permanent component is obtained as

$$\min_{PC} \underbrace{\sum_{t=1}^{T} (y_t - PC_t)^2}_{\text{Variance of CC}} + \lambda \sum_{t=2}^{T-1} \left[(PC_{t+1} - PC_t)^2 + (PC_t - PC_{t-1})^2 \right]. \quad (38)$$

- The bigger is λ, the smoother is the trend.
- In practice the following values are used:
  - $\lambda = 100$ for annual data
  - $\lambda = 1600$ for quarterly data
  - $\lambda = 14400$ for monthly data
- Note that if $\lambda = 0$, it is $PC_t = y_t$.

The Baseline Linear Regression Model - Outline

Exponential Smoothing

Exponential smoothing(ES) is a method to produce short-term forecasts quickly and with sufficient accuracy.

- ES decomposes a time series into a "level" component and an unpredictable residual component.
- Once the level at the end of the estimation sample is obtained, say y_T^L, it is used as a forecast for y_{T+h}, $h > 1$.

Exponential Smoothing

- If y_t is an i.i.d. process with a non-zero mean, we could estimate y_T^L as the sample mean of y.
- If instead y_t is persistent, then the more recent observations should receive a greater weight. Hence, we could use

$$y_T^L = \sum_{t=1}^{T-1} \alpha(1-\alpha)^t y_{T-t},$$

with $0 < \alpha < 1$ and

$$\widetilde{y}_{T+h} = y_T^L. \tag{39}$$

- Since
$$(1-\alpha)y_T^L = \sum_{t=1}^{T-1} \alpha(1-\alpha)^{t+1} y_{T-t-1},$$
we have
$$y_T^L = \alpha y_T + (1-\alpha) y_{T-1}^L,$$
with the starting condition $y_1^L = y_1$.
- The larger α the larger the weight on the most recent observations.
- Note that in the limiting case where $\alpha = 1$, it is $y_T^L = y_T$ and the ES forecasts coincides with that from a RW model for y.

Exponential Smoothing

- A more elaborate model underlies the Holt-Winters procedure

$$\begin{aligned} y_t^L &= a y_t + (1-a)(y_{t-1}^L + T_{t-1}), \\ T_t &= c(y_t^L - y_{t-1}^L) + (1-c)T_{t-1}, \end{aligned} \tag{40}$$

with $0 < a < 1$, and starting conditions $T_2 = y_2 - y_1$ and $y_2^L = y_2$. In this case, we have

$$\widetilde{y}_{T+h} = y_T^L + hT_T. \tag{41}$$

- The coefficients α and a control the smoothness of y_t^L.
- In practice, the smoothing coefficients, and c, are selected by minimizing the in-sample MSFE:

$$\sum_{t=3}^{T-1} (y_t - \widetilde{y}_t)^2.$$

When can the ES forecasts be considered optimal?

- Focusing on the more general case in (41) for which we have:

$$e_t = y_t - \widetilde{y}_{t|t-1|} = y_t - y_{t-1}^L - T_{t-1}. \tag{42}$$

- From (40) and (42) we have:

$$y_t^L - y_{t-1}^L = a(y_t - y_{t-1}^L) + (1-a)T_{t-1} = T_{t-1} + ae_t, \tag{43}$$

- Combining (40) and (43) we have:

$$T_t - T_{t-1} = c(y_t^L - y_{t-1}^L) - cT_{t-1} = cae_t \tag{44}$$

- Using (43) we have

$$\begin{aligned}(y_t^L - y_{t-1}^L) - (y_{t-1}^L - y_{t-2}^L) &= T_{t-1} - T_{t-2} + ae_t - ae_{t-1} \\ &= cae_{t-1} + ae_t - ae_{t-1}\end{aligned}$$

so that

$$(1-L)^2 y_t^L = a\left[1-(1-c)L\right]e_t.$$

- From (44) we have

$$(1-L)^2 T_t = ca(1-L)e_t.$$

- Putting together the expressions for $(1-L)^2 y_t^L$ and $(1-L)^2 T_t$ we obtain
$$(1-L)^2(y_t^L + T_t) = a(1+c)e_t - ae_{t-1}$$
- From (42) $e_t = y_t - y_{t-1}^L - T_{t-1}$, we can write
$$(1-L)^2 y_t = (1-L^2)e_t + a(1+c)e_{t-1} - ae_{t-2}.$$
- Therefore, in conclusion, the Holt-Winter procedure forecast in (41) is optimal (in the MSFE sense) if y_t is an ARIMA(0,2,2).
- Similarly, it can be shown that the forecast in (39) is optimal when y_t is an ARIMA(0,1,1).

Seasonality

- **Seasonality** is a systematic but possibly stochastic intra-year variation in the behavior of an economic variable, related, e.g., to the weather or the calendar and their impact on economic decisions.
    - e.g. sales are higher in December, electricity consumption peaks in the summer
- A common approach in the analysis and forecasting of economic time series is to work with seasonally adjusted variables.

Seasonality

- However, seasonal adjustment procedures can spuriously alter the dynamic behavior of a variable, in particular when the seasonal pattern changes over time.
- Ghysels, Granger, and Siklos (1996) show that seasonal adjusted white noise produces a series that has predictable patterns.
- We will discuss two alternative approaches in this context:
  - Deterministic seasonality
  - Stochastic seasonality

Deterministic Seasonality

Assume that the time series x_t can be decomposed into a deterministic seasonal component, s_t, and an ARMA component, y_t:

$$x_t = s_t + y_t. \tag{45}$$

- Under the additional assumption that t is measured in months and there is a deterministic monthly seasonality, we can write

$$s_t = \sum_{i=1}^{12} \gamma_i D_{it},$$

where D_{it} are seasonal dummy variables taking values 1 in each month i and 0 otherwise, $i = 1, \ldots, 12$.

Deterministic Seasonality

- The seasonal coefficients γ_i can be estimated in the ARMA model for x_t as previously discussed.
- Alternative, we could seasonally adjust the variable x_t by subtracting from each of its values $\widetilde{s}_t$, where

$$\widetilde{s}_t = \sum_{i=1}^{12} \widetilde{\gamma}_i D_{it}, \qquad \widetilde{\gamma}_i = \bar{x}_i,$$

and $\bar{x}_i$ is the sample mean of all month-i observations of x.

Deterministic Seasonality

- We can handle a slow evolution in the shape or amplitude of the seasonal component, by using weighted averages of the month-i observations of x for the construction of $\widetilde{\gamma}_i$.
- For example, also based on the discussion of exponential smoothing in the previous section we could use

$$\widetilde{\gamma}_i(j) = \alpha\widetilde{\gamma}_i(j-1) + (1-\alpha)x_{j,i},$$

  where $\alpha \in [0,1]$, j is measured in years, and $x_{j,i}$ indicates the value of x in month i of year j.
- A similar dummy variable approach can be used for different patterns of seasonality (e.g., quarterly), and also to handle other data irregularities such as working days effects and moving festivals.

Stochastic Seasonality

If the seasonal pattern could be stochastic rather than deterministic, we could extend the ARMA(p,q) specification to take into explicit account the possibility of seasonal dynamics.

- We could use a model such as

$$\phi_1(L)\,\phi_2(L^s)\,y_t = \psi_1(L)\,\psi_2(L^s)\,\varepsilon_t, \tag{46}$$

  where L^s is the seasonal lag operator, so that

$$\begin{aligned} \phi_2(L^s) &= 1 - \phi_{21}L^s - \ldots - \phi_{2p_2}L^{sp_s}, \\ \psi_2(L^s) &= 1 + \psi_{21}L^s + \ldots + \psi_{2\psi_2}L^{sq_s}, \end{aligned}$$

  .

- This model is typically known as a seasonal ARMA model and it can properly represent several seasonal time series.

- One example is the seasonal random walk model:

$$y_t = y_{t-s} + \varepsilon_t,$$

  - year-to-year differences are white noise.

- Another example is the case where p = p_s = q = q_s =1 and s =12:

$$(1-\phi_{11}L)(1-\phi_{21}L^{12})y_t = (1+\psi_{11}L)(1+\psi_{21}L^{12})\varepsilon_t,$$

or

$$(1-\phi_{11}L-\phi_{21}L^{12}+\phi_{11}\phi_{21}L^{13})y_t = (1+\psi_{11}L+\psi_{21}L^{12}+\psi_{11}\psi_{21}L^{13})\varepsilon_t.$$

- This can also be considered as an ARMA(13,13), though with a set of zero restrictions on the coefficients.

- Since a seasonal ARMA is just a constrained version of a general ARMA model, the same tools for specification, estimation and diagnostic testing that we have seen for the ARMA case can be applied in this context.
- Finally, some of the roots of the seasonal AR polynomial $\phi_2(L^s)$ could be equal to one as in the seasonal random walk model. In the general case, these are known as seasonal unit roots.

The Baseline Linear Regression Model - Outline

Examples With Simulated Data

- We use standard Box–Jenkins procedure to specify and estimate several types of ARMA models. Once the appropriate ARMA specification is found, it is used for an out-of-sample forecasting exercise.
- The data set contains 600 simulated observations.
- In order to avoid dependence on the starting values, the first 100 observations are discarded.

Stationary ARMA Process

We consider three stationary ARMA process:

- y_1, with an AR(2) DGP
- y_2, with an MA(2) DGP
- y_3, with an ARMA(2,2) DGP

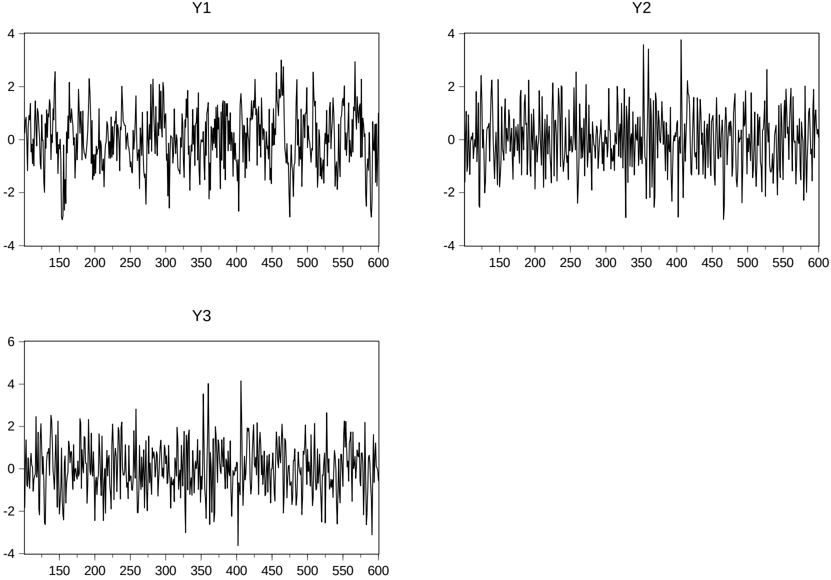
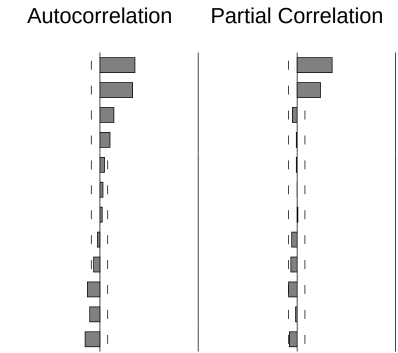


Figure 1: *Stationary ARMA process y_1 through y_3, sample 101 - 600*

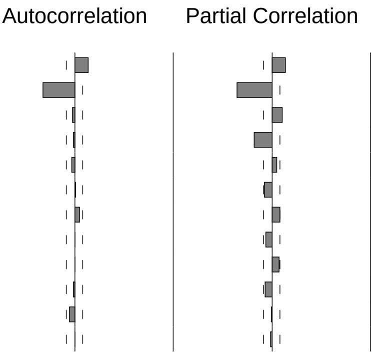
- Note from Figure 1 that the dynamic behavior of the variables looks similar although they have been generated from different DGPs.
- Figures 2 - 4 show the (partial) correlograms of the three series for the entire sample under consideration.

Correlogram for y_1



	AC	PAC
1	0.370	0.370
2	0.350	0.246
3	0.149	-0.049
4	0.115	-0.002
5	0.054	-0.001
6	0.040	0.003
7	0.029	0.010
8	-0.024	-0.054
9	-0.058	-0.058
10	-0.124	-0.089
11	-0.099	-0.011
12	-0.154	-0.074

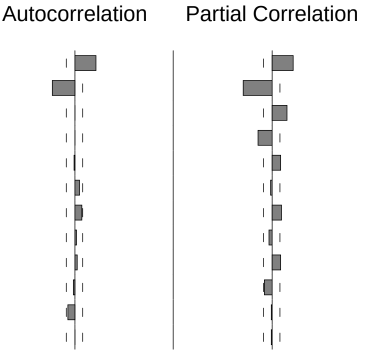
Figure 2: *AC and PAC functions for y_1 entire sample (101 -600)*



	AC	PAC
1	0.145	0.145
2	-0.335	-0.364
3	-0.022	0.114
4	-0.014	-0.184
5	-0.029	0.054
6	0.009	-0.074
7	0.055	0.089
8	0.004	-0.060
9	0.003	0.079
10	-0.010	-0.067
11	-0.056	-0.001
12	0.005	-0.010

Figure 3: *AC and PAC functions for y_2 entire sample (101 - 600)*

Correlogram for y_3



	AC	PAC
1	0.227	0.227
2	-0.234	-0.301
3	0.002	0.161
4	0.001	-0.140
5	-0.001	0.097
6	0.052	-0.013
7	0.079	0.105
8	0.020	-0.029
9	0.032	0.097
10	-0.014	-0.082
11	-0.072	-0.001
12	0.008	-0.003

Figure 4: *AC and PAC functions for y_3 entire sample (101 - 600)*

- Figure 2, y_1 has an exponentially decaying autocorrelation function and significant peaks at lags one and two in the PAC, strongly supporting an AR(2) specification.
- Figure 3 has an oscillating PAC function while the AC displays a significant peak at lag two, suggesting that y_2 could in fact follow a MA(2) process.
- From Figure 4 it is harder to make a proposal for a model for y_3 based on the reported AC and PAC functions. However, the presence of significant peaks in both of them suggests a combined ARMA specification.

Let's use BIC to select an appropriate ARIMA specification and compare with the AC and PAC analysis.

- Below is the BIC for various specifications for y_1.

AR / MA	0
0	3.042114
1	2.909087
2	**2.860312**
3	2.871777
4	2.885768
5	2.896720
6	2.910012

Table 1: *The ARIMA selection with BIC for* y_1

AR / MA	0	1	2	3
0	3.038216	2.965412	**2.817112**	2.820006
1	3.024794	2.871028	2.819972	2.823934
2	2.886401	2.836022	2.823942	2.827933
3	2.877178	2.830599	2.827901	2.823809

Table 2: *The ARIMA selection with BIC for* y_2

AR / MA	0	1	2	3
0	3.187754	3.059214	2.981900	2.985839
1	3.142723	2.995391	2.985840	2.988778
2	3.051139	2.987988	**2.981839**	2.991453
3	3.028714	2.989539	2.993529	2.987754

Table 3: *The ARIMA selection with BIC for* y_3

- we see that the procedure always selects the true DGPs when the entire sample is taken as a reference, as shown in Tables 1 - 3 for BIC.
- Next, we will estimate the specifications selected by BIC.
- Finally, we will illustrate the forecasting performance of the selected models.

Estimation: Full sample analysis

Tables 4 through 6 report the estimation results of the three models suggested by the ARIMA selection procedure;

- y_1 is modeled as an AR(2)
- y_2 is modeled as an MA(2)
- y_3 is modeled as an ARMA(2,2)

Note that the selected AR and MA terms are always significant and Durbin-Watson statistics are always pretty close to the value of 2, signaling almost with noise residuals.

Estimation results for y_1

Dep Var: Y1				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.042014	0.093569	0.449014	0.653
AR(1)	0.279133	0.043545	6.410269	0.000
AR(2)	0.245699	0.043530	5.644410	0.000
R-squared	0.189038	Mean dep var		0.03953
Adjusted R-squared	0.185774	S.D. dep var		1.10175
S.E. of regression	0.994163	Akaike IC		2.83215
Sum squared resid	491.2152	Schwarz IC		2.85743
Log likelihood	-705.0378	Hannan-Quinn		2.84207
F-statistic	57.92603	DW stat		1.96986
Prob(F-statistic)	0.000000			

Table 4: *Estimation results for* y_1

Estimation results for y_2

Dep Var: Y2 Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.009993	0.038195	-0.261621	0.793
MA(1)	0.281535	0.040840	6.893551	0.000
MA(2)	-0.414625	0.040851	-10.14959	0.000
R-squared	0.209487	Mean dep var		-0.01033
Adjusted R-squared	0.206306	S.D. dep var		1.10425
S.E. of regression	0.983775	Akaike IC		2.81114
Sum squared resid	481.0032	Schwarz IC		2.83643
Log likelihood	-699.7857	Hannan-Quinn		2.82106
F-statistic	65.85277	DW stat		1.96942
Prob(F-statistic)	0.000000			

Table 5: *Estimation results for* y_2

Dep Var: Y3 Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011417	0.054939	0.207816	0.835
AR(1)	-0.193760	0.283357	-0.683803	0.494
AR(2)	-0.076268	0.132683	-0.574815	0.565
MA(1)	0.594037	0.284018	2.091550	0.037
MA(2)	-0.135818	0.241172	-0.563158	0.573
R-squared	0.198335	Mean dep var		0.01040
Adjusted R-squared	0.191857	S.D. dep var		1.18998
S.E. of regression	1.069757	Akaike IC		2.98269
Sum squared resid	566.4681	Schwarz IC		3.02483
Log likelihood	-740.672	Hannan-Quinn		2.99922
F-statistic	30.61621	DW stat		1.99818
Prob(F-statistic)	0.000000			

Table 6: *Estimation results for* y_3

- Diagnostics appear in Tables 7 though 9, showing some routine statistics for the residuals coming from the estimated models for y_1, y_2 and y_3.

Diagnostic tests on the AR(2) model for y_1

Heteroskedasticity Test: ARCH			
F-statistic	0.009042	F(1,497)	0.924
Obs*R-squared	0.009078	Chi-Sq(1)	0.924
Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.628107	F(2,495)	0.534
Obs*R-squared	1.265691	Chi-Sq(2)	0.531
Heteroskedasticity Test: White			
F-statistic	0.776695	F(5,494)	0.566
Obs*R-squared	3.899984	Chi-Sq(5)	0.563
Scaled explained SS	3.132991	Chi-Sq(5)	0.679
Jarque-Bera normality Test:	3.516553		0.172

Table 7: *Diagnostic tests on the AR(2) model for* y_1

Diagnostic tests on the MA(2) model for y_2

Heteroskedasticity Test: ARCH			
F-statistic	0.169720	F(1,497)	0.680
Obs*R-squared	0.170345	Chi-Sq(1)	0.679
Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.281619	F(2,495)	0.754
Obs*R-squared	0.568031	Chi-Sq(2)	0.752
Heteroskedasticity Test: White			
F-statistic	0.776314	F(9,490)	0.638
Obs*R-squared	7.029190	Chi-Sq(9)	0.634
Scaled explained SS	7.486218	Chi-Sq(9)	0.586
Jarque-Bera normality Test:	0.945034	Prob.	0.623

Table 8: *Diagnostic tests on the MA(2) model for* y_2

Diagnostic tests on the ARMA(2,2) model for y_3

Heteroskedasticity Test: ARCH			
F-statistic	0.087968	F(1,497)	0.766
Obs*R-squared	0.088306	Chi-Sq.(1)	0.766
Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	3.380311	F(2,495)	0.034
Obs*R-squared	6.736898	Chi-Sq.(2)	0.034
Heteroskedasticity Test: White			
F-statistic	1.475130	F(9,490)	0.154
Obs*R-squared	13.18975	Chi-Sq.(9)	0.154
Scaled explained SS	13.83836	Chi-Sq.(9)	0.128
Jarque-Bera normality Test:	0.504597	Prob.	0.777

Table 9: *Diagnostic tests on the ARMA(2,2) model for* y_3

Residuals

- For all the three models, the actual vs fitted and residuals (see Figures 5 through 9) show a good fit for the selected models.
- The residuals display frequent sign changes, although there might be a few outliers.

Actual vs fitted residuals

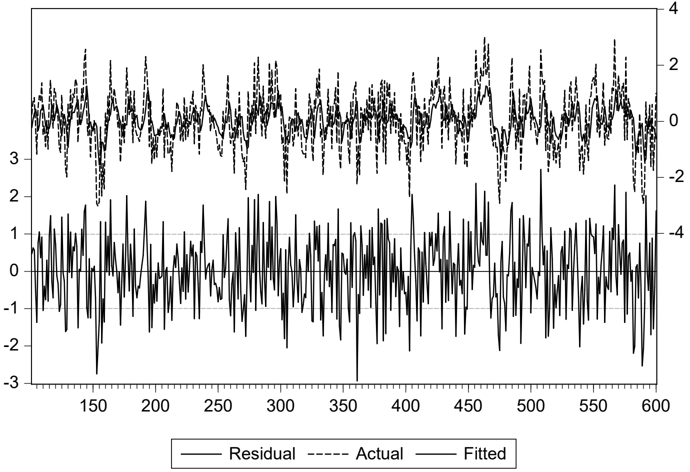
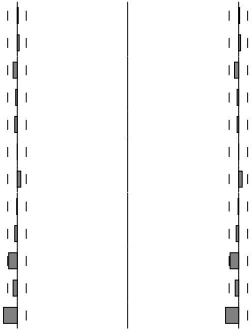


Figure 5: *Actual vs fitted residuals*



Autocorrelation	Partial Correlation		AC	PAC
		1	0.012	0.012
		2	0.019	0.019
		3	-0.040	-0.041
		4	-0.011	-0.011
		5	-0.018	-0.016
		6	0.007	0.007
		7	0.038	0.038
		8	-0.001	-0.004
		9	-0.021	-0.023
		10	-0.078	-0.075
		11	-0.035	-0.032
		12	-0.124	-0.122

Figure 6: *Correlogram of the residuals for the AR(2) model for* y_1

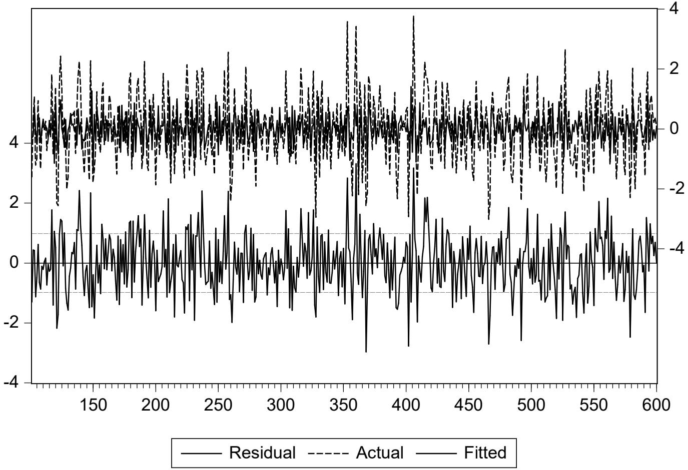
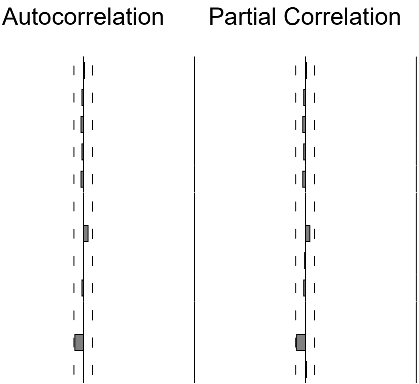


Figure 7: *Actual vs fitted residuals for* y_2



	AC	PAC
1	0.014	0.014
2	-0.009	-0.009
3	-0.023	-0.023
4	-0.011	-0.011
5	-0.022	-0.022
6	0.002	0.002
7	0.044	0.043
8	0.001	-0.002
9	-0.010	-0.010
10	0.005	0.007
11	-0.080	-0.080
12	0.005	0.009

Figure 8: *Correlogram of the residuals for the MA(2) model for* y_2

Actual vs fitted residuals for y_3

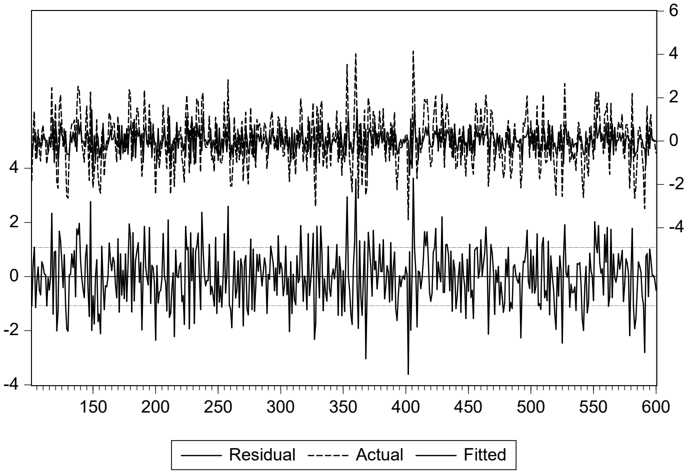
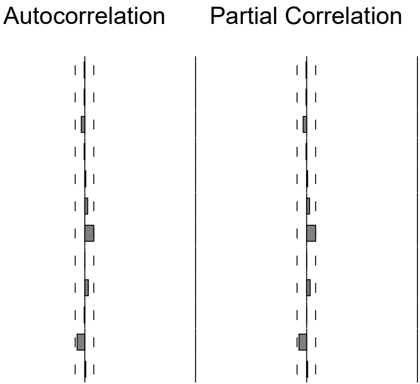


Figure 9: *Actual vs fitted residuals for* y_3

Correlogram of the residuals for the model for y_3



	AC	PAC
1	-0.000	-0.000
2	-0.000	-0.000
3	-0.025	-0.025
4	-0.004	-0.004
5	0.011	0.011
6	0.030	0.029
7	0.083	0.083
8	0.002	0.002
9	0.035	0.037
10	-0.002	0.002
11	-0.072	-0.073
12	0.013	0.011

Figure 10: *Correlogram of the residuals for the ARMA(2,2) model for* y_3

Having a simulated dataset allows us to answer also another interesting question: What if the researcher “gets it wrong” and estimates the following models:

- AR(1) for y_1
- MA(1) for y_2
- ARMA(1,1) for y_3

For y_1 estimating an AR(1) yields a reasonable good fit with only a slight decrease in the

Model Misspecification

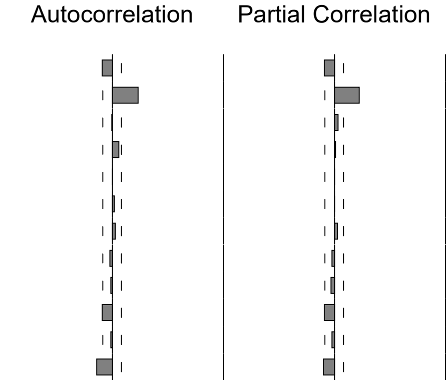
- For y_1 estimating an AR(1) yields a reasonable good fit with only a slight decrease in the in the adjusted R^2 to 14% but no major signs of misspecification (see next slide).
- Nevertheless, the correlogram of the residuals of the AR(1) question features a relevant peak exactly at lag 2, as the AR(2) term was omitted in the estimation.

Estimation results for the mis-specified model for y_1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025639	0.045844	0.559268	0.576
AR(1)	0.370487	0.041659	8.893349	0.000
R-squared	0.137052	Mean dep var		0.03953
Adjusted R-squared	0.135319	S.D. dep var		1.10175
S.E. of regression	1.024503	Akaike IC		2.89028
Sum squared resid	522.7037	Schwarz IC		2.90714
Log likelihood	-720.570	Hannan-Quinn		2.89689
F-statistic	79.09166	DW stat		2.17887
Prob(F-statistic)	0.000000			

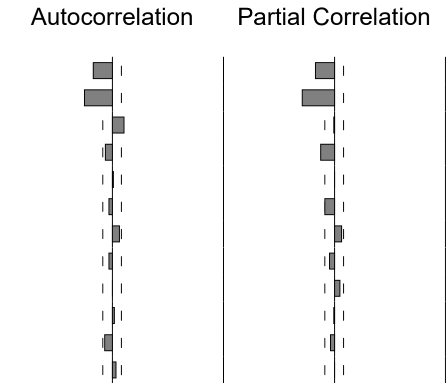
Table 10: *Estimation results for the mis-specified model for* y_1

Correlograms for Mis-specified Model for y_1



	AC	PAC
1	-0.091	-0.091
2	0.238	0.231
3	-0.004	0.036
4	0.065	0.013
5	0.003	0.002
6	0.018	0.000
7	0.031	0.031
8	-0.020	-0.022
9	-0.012	-0.031
10	-0.096	-0.098
11	-0.013	-0.021
12	-0.140	-0.105

Figure 11: *Correlograms for mis-specified model for* y_1



	AC	PAC
1	-0.179	-0.179
2	-0.259	-0.300
3	0.112	-0.003
4	-0.062	-0.130
5	0.011	0.003
6	-0.025	-0.084
7	0.067	0.071
8	-0.027	-0.041
9	0.006	0.051
10	0.018	-0.006
11	-0.071	-0.038
12	0.038	0.005

Figure 12: *Correlograms for mis-specified model for* y_2

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.036	-0.036
		2	-0.101	-0.102
		3	-0.026	-0.034
		4	0.022	0.009
		5	-0.018	-0.023
		6	0.044	0.046
		7	0.061	0.063
		8	0.006	0.019
		9	0.025	0.043
		10	0.005	0.012
		11	-0.078	-0.072
		12	0.029	0.026

Figure 13: *Correlograms for mis-specified model for* y_3

Subsample Analysis

We now will apply the ARIMA selection procedure to two smaller samples: observations 101 to 200 and 201 to 300

	Subsample 101-200		Subsample 201-300	
	AIC	BIC	AIC	BIC
$y1$	AR(2)	AR(1)	AR(2)	AR(2)
$y2$	MA(2)	MA(2)	MA(2)	MA(10)
$y3$	ARMA(2,2)	ARMA(2,2)	ARMA(8,8)	ARMA(3,3)

- As we can see, in shorter samples the criteria can select models rather different from the DGPs.
- This is not necessarily a bad outcome, as long as the residuals of the models satisfy the usual assumptions and the forecasting performance remains satisfactory.

For illustrative purposes here we chose to estimate the three models from observation 100 to 499 and to forecast the last 100 observations of the sample, although 100 periods ahead is quite a large forecast horizon.

- The static forecast is simply the one-step ahead forecast.
- The dynamic forecasting method calculates multi-steps ahead forecasts, from 1- up to 100-period ahead.

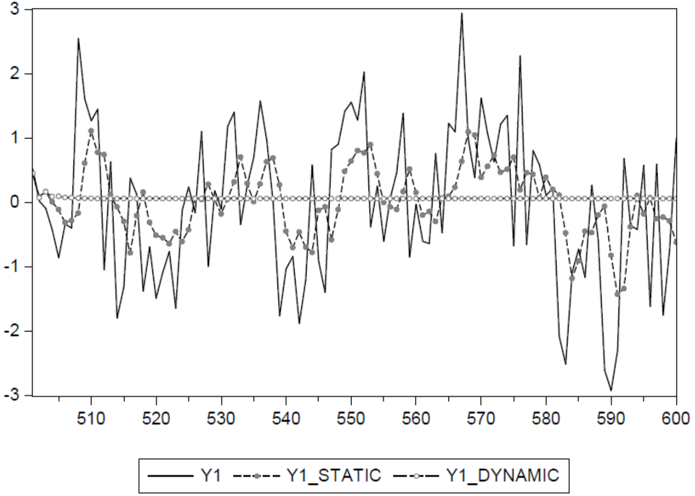


Figure 14: *Actual values and static vs dynamic forecasts for* y_1 *in the sample 501 - 600*

Actual values and forecasts for y_2

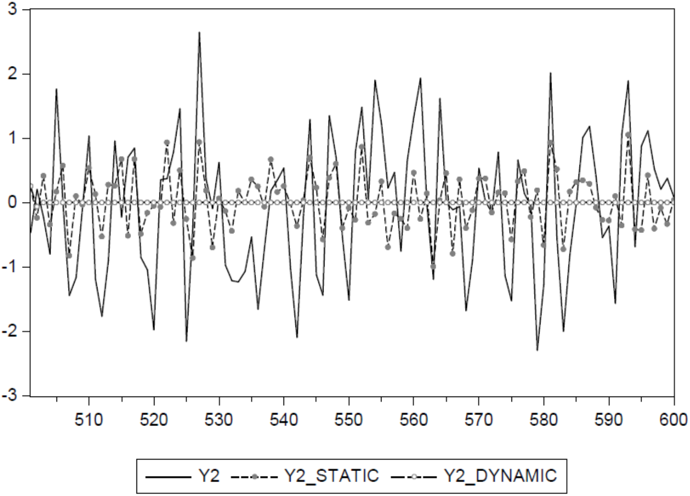


Figure 15: *Actual values and static vs dynamic forecasts for* y_2 *in the sample 501 - 600*

Actual values and forecasts for y_3

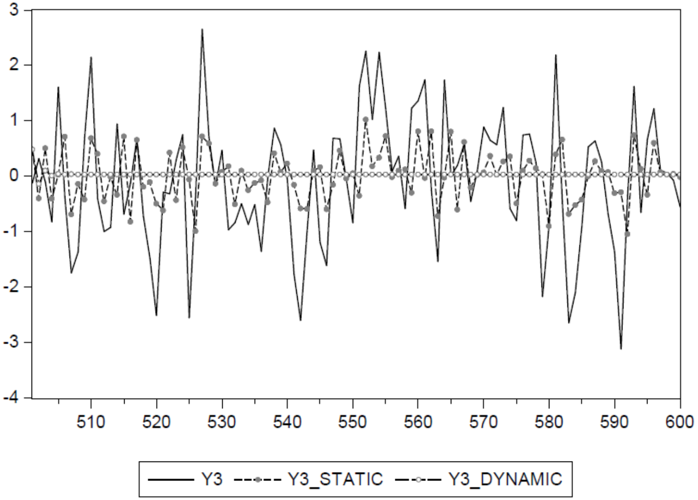


Figure 16: *Actual values and static vs dynamic forecasts for y_3 in the sample 501 - 600*

The Baseline Linear Regression Model - Outline

Modeling and forecasting the US federal funds rate

The aim of this empirical application is to formulate an ARMA model suitable to describe the dynamics of the effective federal funds rate r in the United States.

- Monthly dataset spanning from January 1985 until the end of 2012 from the Federal Reserve Economic Data (FRED) website.
- We will first analyze the time frame going from 1985 to 2002, using the period from 2003 to 2006 for an out-of-sample forecasting exercise prior the global financial crisis.
- Then, we will enlarge the estimation sample to the start of the crisis, i.e., in August 2007, and we will see whether our model is able to forecast the steep fall that the federal funds rate displayed from 2008 until the end of the sample.

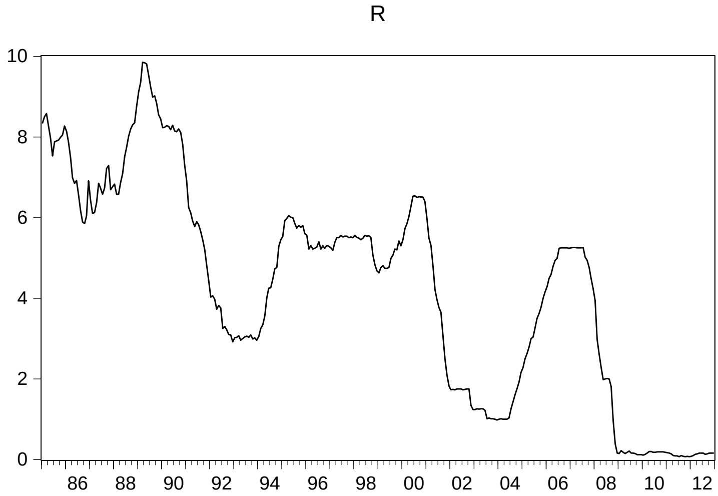


Figure 17: *The effective US federal funds rate 1985 - 2012*

- The Augmented Dickey-Fuller test can help us investigate the dynamic properties of the series more thoroughly.
- Based on the results in the following slide, we cannot reject the null of a unit root at conventional significance levels.
- However, when we perform the same test on the differenced variable Dr instead, there is strong evidence in favor of stationarity, suggesting that a specification in differences for r is preferred.

The Augmented Dickey-Fuller test

ADF Test for r		t-Statistic	Prob.*
		-0.912779	0.782
Test critical values:	1% level	-3.460884	
	5% level	-2.874868	
	10% level	-2.573951	
ADF Test for Dr		**t-Statistic**	**Prob.***
		-8.810261	0.000
Test critical values:	1% level	-2.575813	
	5% level	-1.942317	
	10% level	-1.615712	

Table 11: *The Augmented Dickey-Fuller test*

Lag Specification

- BIC suggests a ARMA(10,1) for Dr (see next slide).
- However, AIC suggest ARMA(5,2) for Dr.
- We started specifying an ARMA(10,2) model for Dr, and then we noticed that some AR lags could be conveniently eliminated. At the end of this general-to-specific specification procedure we ended up with a parsimonious ARMA(5,2) model reported in Table 13.

ARIMA selection with BIC for Dr

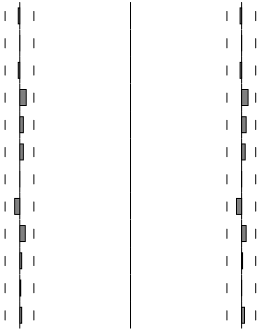
AR / MA	0	1	2
0	-0.276287	-0.514755	-0.544955
1	-0.596637	-0.611789	-0.627400
2	-0.600159	-0.622702	-0.622388
3	-0.615346	-0.621617	-0.616286
4	-0.611976	-0.615451	-0.610219
5	-0.619612	-0.615223	-0.621251
6	-0.616315	-0.612577	-0.621058
7	-0.613225	-0.607978	-0.621052
8	-0.611476	-0.607423	-0.611349
9	-0.614132	-0.610542	-0.609949
10	-0.615026	**-0.650287**	-0.644754
11	-0.615257	-0.645099	-0.641540
12	-0.609648	-0.640053	-0.638794

Table 12: *ARIMA selection with BIC for Dr*

Parameter estimates ARMA(5,2) model for Dr

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.169824	0.063523	2.673423	0.008
AR(2)	-0.799708	0.043098	-18.55554	0.000
AR(3)	0.515466	0.076887	6.704226	0.000
AR(5)	0.205128	0.066405	3.089048	0.002
MA(1)	0.262565	0.013437	19.54028	0.000
MA(2)	0.976636	0.009565	102.1080	0.000
R-squared	0.282454	Mean dep var		-0.02995
Adjusted R-squared	0.264867	S.D. dep var		0.22866
S.E. of regression	0.196060	Akaike IC		-0.39263
Sum squared resid	7.841698	Schwarz IC		-0.29700
Log likelihood	47.22639	Hannan-Quinn		-0.35397
DW stat	2.004751			

Table 13: *Parameter estimates ARMA(5,2) model for Dr*



Autocorrelation	Partial Correlation		AC	PAC
		1	-0.016	-0.016
		2	0.005	0.005
		3	-0.014	-0.014
		4	0.064	0.064
		5	0.040	0.042
		6	0.035	0.036
		7	0.000	0.003
		8	-0.044	-0.047
		9	0.049	0.044
		10	0.020	0.016
		11	0.011	0.007
		12	0.021	0.027

Figure 18: *Correlogram of residuals from ARMA(5,2) for Dr*

Breusch-Godfrey			
Serial Correlation LM Test:			
F-statistic	0.090390	F(2,202)	0.913
Obs*R-squared	0.187771	Chi-Sq(2)	0.910
Heteroskedasticity Test: ARCH			
F-statistic	35.80092	F(1,207)	0.000
Obs*R-squared	30.81699	Chi-Sq(1)	0.000
Heteroskedasticity Test: White			
F-statistic	4.645695	F(21,188)	0.000
Obs*R-squared	71.74515	Chi-Sq(21)	0.000

Table 14: *Diagnostic tests on the residuals of ARMA(5,2) model for Dr*

Forecasting

We now want to produce one-step ahead forecasts for r in the period 2003 - 2006.

- The first option is to construct forecasts for Dr and then cumulate them with the starting value for r, namely,

$$\widehat{r}_{T+1} = r_T + \widehat{Dr}_{T+1}.$$

  - When we use this method, we add to the forecast series the suffix "differences," in order to clarify that the forecast for r_{T+1} has been obtained indirectly from the model specified in differences.
- As an alternative, we can specify a model for the level of the interest rate r and produce a forecast for r_{T+1} directly.

Which model is suitable to represent r in levels?

- The unit root test indicated that there was evidence for non-stationarity, and the correlogram showed a very persistent AR type of behavior.
- It turns out that an AR(11) specification, with many non-significant intermediate lags removed, is able to capture almost all the variability of the series.
- From The estimation output in Table 15, we note the very high R^2, which is not surprising for a specification for the levels of an I(1) variable.
- The diagnostic tests on the residuals reported in Table 16 confirm the likely correctly specified dynamics.

Parameter estimates restricted AR(11) model for r

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.410725	0.065970	21.38418	0.000
AR(2)	-0.385824	0.071123	-5.424708	0.000
AR(11)	-0.028306	0.011827	-2.393426	0.017
R-squared	0.989310	Mean dep var		5.51282
Adjusted R-squared	0.989204	S.D. dep var		1.92831
S.E. of regression	0.200362	Akaike IC		-0.36285
Sum squared resid	8.109257	Schwarz IC		-0.31422
Log likelihood	40.19299	Hannan-Quinn		-0.34318
DW stat	2.004220			

Table 15: *Parameter estimates restricted AR(11) model for r*

Diagnostic tests on the residuals of AR(11) model for r

Breusch-Godfrey			
Serial Correlation LM Test:			
F-statistic	0.660035	F(2,200)	0.518
Obs*R-squared	1.344200	Chi-Sq(2)	0.510
Heteroskedasticity Test: ARCH			
F-statistic	26.24542	F(1,202)	0.000
Obs*R-squared	23.45749	Chi-Sq(1)	0.000
Heteroskedasticity Test: White			
F-statistic	11.92362	F(6,198)	0.000
Obs*R-squared	54.41108	Chi-Sq(6)	0.000

Table 16: *Diagnostic tests on the residuals of AR(11) model for r*

Forecasts

- We can now compare the forecasting performance of the models specified in differences and in levels.
- Figure 19 shows dynamic or multi-steps ahead forecasts.
- Figure 20 shows static or one-step ahead forecasts.
- Table 17 complements the analysis by showing the standard forecast evaluation criteria, the RMSFE and the MAFE, for the one-step ahead forecasts.

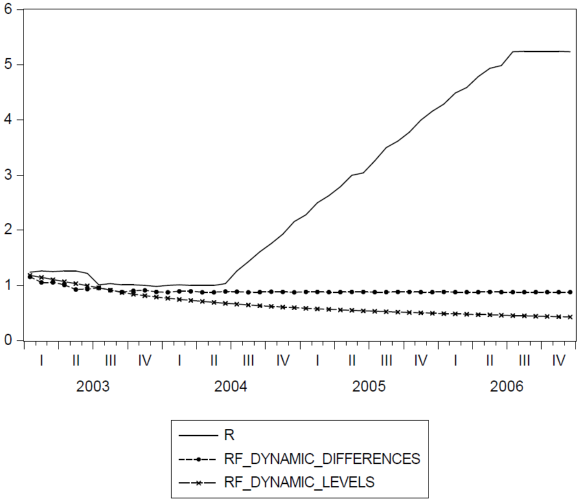


Figure 19: *The actual values of r against the forecasted series: dynamic or multi-steps ahead forecasts for the period 2003 - 2006*

Static or One-step Ahead Forecast

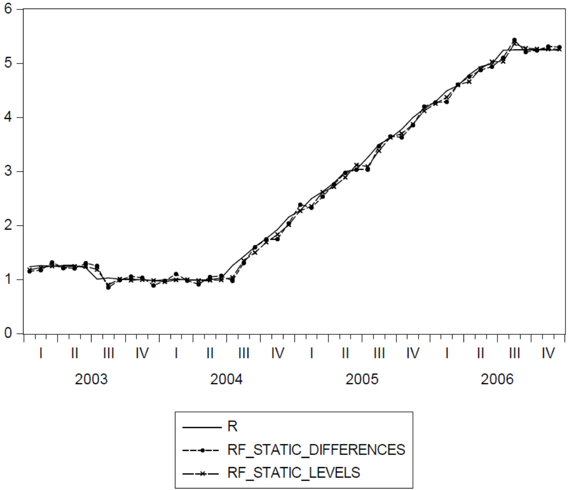


Figure 20: *The actual values of* r *against the forecasted series: static or one-step ahead forecast, for the period 2003 - 2006.*

Forecast Evaluations for the One-step Ahead Forecasts

	RMSFE	MAFE
Levels	0.0868	0.0637
Differences	0.1075	0.0834

Table 17: *Forecast evaluations for the one-step ahead forecasts*

The Crisis Period

Will the one-step predictive capacity of the above-selected models be any different if we enlarge our to include the financial crisis period?

- Tables 18 and 20 clearly show that the models' fit remains quite similar also using the enlarged estimation sample.
- Figures 21, 22, and Table 19 report detailed forecast evaluation criteria.
- We also look at 2-step ahead forecast in Figure 23.

Parameter estimates ARMA(5,2) for Dr with enlarged sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.309924	0.179822	1.723511	0.086
AR(2)	-0.529328	0.201804	-2.622974	0.009
AR(3)	0.385577	0.096964	3.976501	0.000
AR(5)	0.227198	0.056587	4.015022	0.000
MA(1)	0.099835	0.180357	0.553539	0.580
MA(2)	0.613386	0.175845	3.488216	0.000
R-squared	0.290327	Mean dep var		-0.00856
Adjusted R-squared	0.276627	S.D. dep var		0.21258
S.E. of regression	0.180808	Akaike IC		-0.56037
Sum squared resid	8.467142	Schwarz IC		-0.47932
Log likelihood	80.24991	Hannan-Quinn		-0.52781
DW stat	2.007645			

Table 18: *Parameter estimates ARMA(5,2) for Dr with enlarged sample*

Forecast Evaluations for the One-step Ahead Forecasts

	RMSFE	MAFE
Levels	0.1491	0.0694
Differences	0.1552	0.0703

Table 19: *Forecast evaluations for the one-step ahead forecasts*

Parameter estimates AR(11) model for r with enlarged sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.417957	0.058500	24.23864	0.000
AR(2)	-0.388422	0.063316	-6.134685	0.000
AR(11)	-0.032273	0.010137	-3.183665	0.001
R-squared	0.992941	Mean dep var		4.97984
Adjusted R-squared	0.992886	S.D. dep var		2.15119
S.E. of regression	0.181445	Akaike IC		-0.56425
Sum squared resid	8.461069	Schwarz IC		-0.52316
Log likelihood	76.35278	Hannan-Quinn		-0.54773
DW stat	2.009223			

Table 20: *Parameter estimates AR(11) model for r with enlarged sample*

Actual vs 1-step Forecast

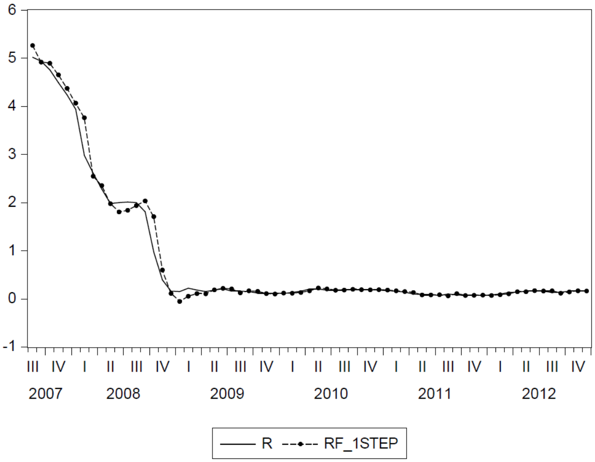


Figure 21: *Actual vs 1-step forecasted r series, period 2007 - 2012*

Actual vs h-step Ahead Forecast

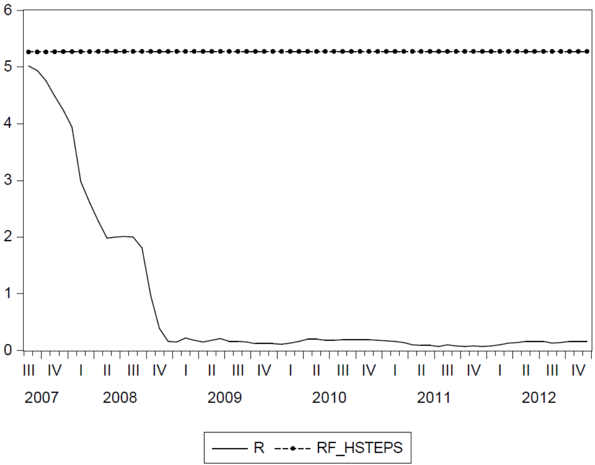


Figure 22: *Actual vs h-step forecasted r series, period 2007 - 2012*

Actual vs 2-steps Ahead Forecast

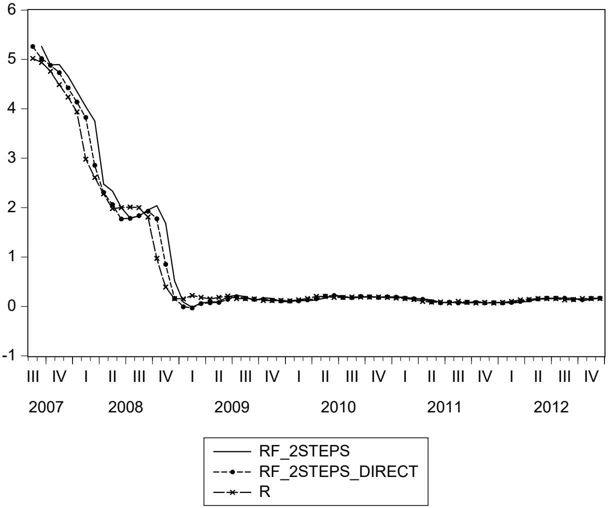


Figure 23: *Actual vs 2-steps ahead forecasted r series, period 2007 - 2012*

	One-step ahead	Two-steps iterated	ahead direct
RMSFE	0.1552	0.3160	0.1813
MAFE	0.0703	0.1461	0.1813

Table 21: *Forecast evaluation*

Change in Real Private Inventories

As a second empirical example, we consider the quarterly time series of the change in real private inventories ($rcpi$).

- Data for the United States, for the period 1985 - 2012, downloaded again from the FRED database.
- The series seems stationary around its mean while its variance seems to increase towards the end of the sample.
- We begin our analysis by inspecting the AC and PAC.

rcpi in the US for 2007-2012

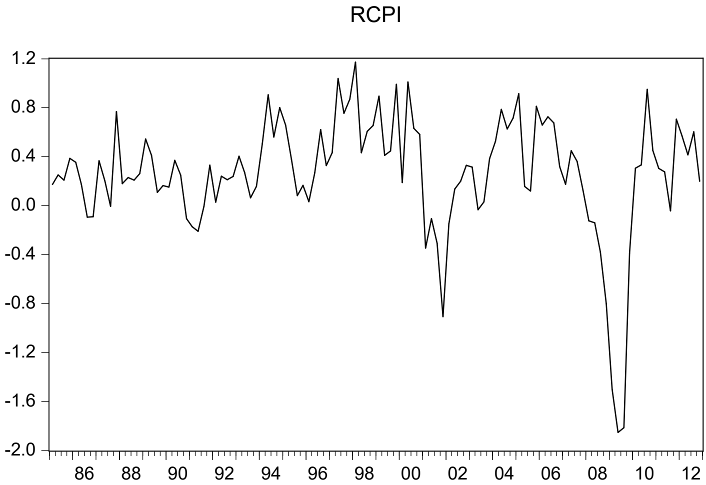
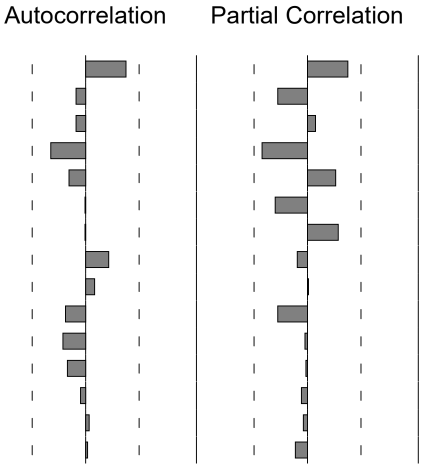


Figure 24: *The series of real changes in private inventories for the US for 2007 - 2012*

Correlogram of *rcpi*



	AC	PAC
1	0.385	0.385
2	-0.083	-0.272
3	-0.087	0.075
4	-0.320	-0.421
5	-0.152	0.263
6	-0.003	-0.303
7	-0.002	0.290
8	0.220	-0.093
9	0.089	0.016
10	-0.186	-0.272
11	-0.206	-0.021
12	-0.165	-0.009
13	-0.048	-0.054
14	0.039	-0.035
15	0.020	-0.112

Figure 25: *Correlogram of rcpi*

Change in Real Private Inventories

- The ADF confirms the time series is stationary.
- AIC- and BIC based results both suggest an ARMA(3,2).
- Estimation and diagnostics follow.

Augmented Dickey-Fuller Test for rcpi

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.9513	0.002
Test critical values:	1% level	-3.4970	
	5% level	-2.8906	
	10% level	-2.5823	

Table 22: *Augmented Dickey-Fuller test for rcpi*

Parameter estimates ARMA(3,2) model for *rcpi*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.311431	0.094479	3.296302	0.001
AR(1)	0.596686	0.113719	5.247034	0.000
AR(2)	-0.712288	0.098247	-7.249984	0.000
AR(3)	0.451393	0.106507	4.238135	0.000
MA(1)	-0.154118	0.034152	-4.512663	0.000
MA(2)	0.967220	0.020041	48.26132	0.000
R-squared	0.433316	Mean dep var		0.31330
Adjusted R-squared	0.388342	S.D. dep var		0.36738
S.E. of regression	0.287329	Akaike IC		0.42656
Sum squared resid	5.201136	Schwarz IC		0.62083
Log likelihood	-8.71634	Hannan-Quinn		0.50363
F-statistic	9.634632	DW stat		2.01732
Prob(F-statistic)	0.000001			

Table 23: *Parameter estimates ARMA(3,2) model for rcpi*

Actual vs fitted and residuals from ARMA(3,2) for $rcpi$, 1986 - 2002

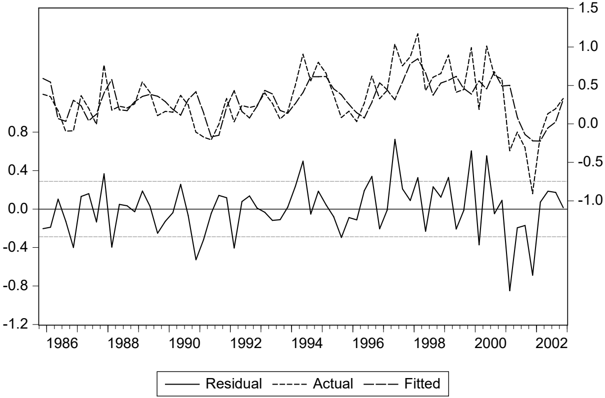
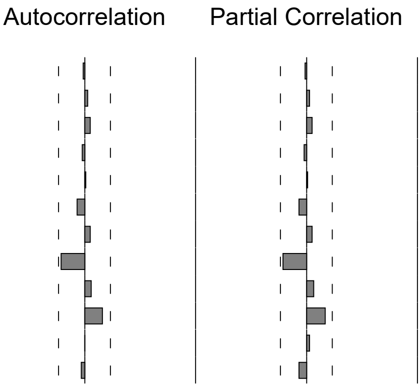


Figure 26: *Actual vs fitted and residuals from ARMA(3,2) for $rcpi$, 1986 - 2002*

Correlogram of the residuals from ARMA(3,2)



	AC	PAC
1	-0.013	-0.013
2	0.029	0.029
3	0.054	0.055
4	-0.017	-0.017
5	0.015	0.012
6	-0.071	-0.073
7	0.054	0.053
8	-0.215	-0.214
9	0.064	0.073
10	0.165	0.175
11	0.005	0.031
12	-0.031	-0.070

Figure 27: *Correlogram of the residuals from ARMA(3,2)*

Diagnostic tests on the residuals from ARMA(3,2)

Breusch-Godfrey			
Serial Correlation LM Test:			
F-statistic	0.054017	F(2,61)	0.947
Obs*R-squared	0.121987	Chi-Sq(2)	0.940
Heteroskedasticity Test: ARCH			
F-statistic	0.625783	F(1,66)	0.431
Obs*R-squared	0.638691	Chi-Sq(1)	0.424
Heteroskedasticity Test: White			
F-statistic	0.443385	F(27,41)	0.985
Obs*R-squared	15.59381	Chi-Sq(27)	0.960
Scaled explained SS	19.89035	Chi-Sq(27)	0.835

Table 24: *Diagnostic tests on the residuals from ARMA(3,2)*

- We will compare both one-step ahead/static, multi-steps ahead/dynamic forecasts and two steps ahead forecast.
- The quality of these forecasts is not very high.

The one-step ahead and the h-steps ahead forecast series against the actuals for $rcpi$, 1986 - 2002

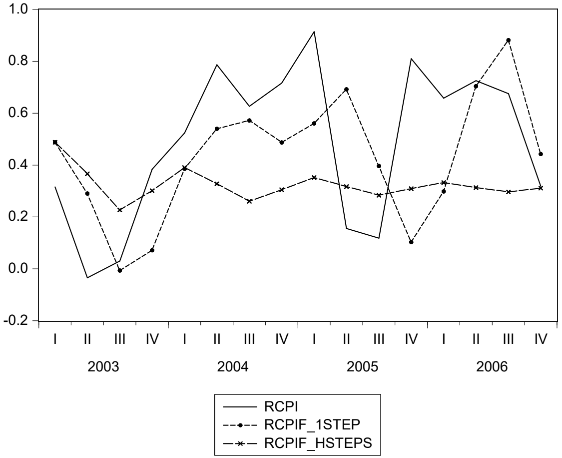


Figure 28: *The one-step ahead and the h-steps ahead forecast series against the actuals for* $rcpi$, *1986 - 2002*

The two-steps ahead forecasts against the actuals for $rcpi$, 1986 - 2002

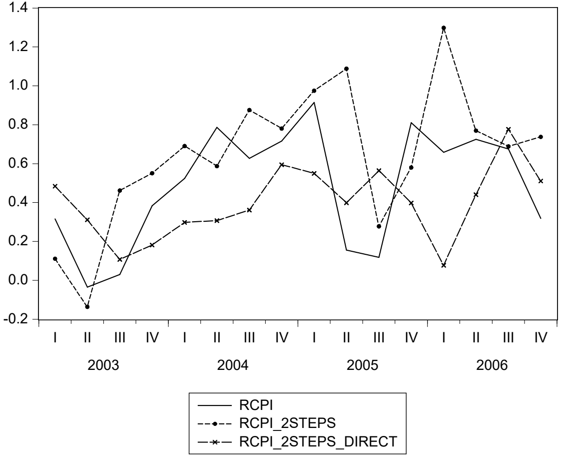


Figure 29: *The two-steps ahead forecasts against the actuals for* $rcpi$, *1986 - 2002*

	1-step ahead	2-steps ahead, iterated	2-steps ahead, direct	1-to-x steps ahead
RMSFE	0.3110	0.5800	0.3152	0.3362
MAFE	0.2563	0.4393	0.2822	0.2961

Table 25: *Forecast evaluation*

The Crisis Period

We now assess whether an ARMA model can produce reliable forecasts also during the global financial crisis period.

- We now include the crisis period in the estimation sample, which spans from 1985Q1 to 2009Q4, and we produce forecasts for 2010Q1 - 2012Q4.
- We use a dummy variable equal to 1 in the period 2007Q4 - 2009Q4 and equal to 0 elsewhere.
- The selected most parsimonious specification is reported in Table 26.
- A seventh-order AR term had to be included, as otherwise the correlogram indicated significant serial autocorrelation. The term is only mildly significant if multiplied by the dummy, but it substantially improves the overall model fit.

Estimation results for the AR(7) model for *rcpi*

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.144059	0.048902	2.945845	0.004
C(2)	-0.567565	0.410887	-1.381317	0.170
C(3)	0.465738	0.109646	4.247637	0.000
C(4)	2.409279	0.546987	4.404637	0.000
C(5)	-2.893790	0.726124	-3.985256	0.000
C(6)	0.134667	0.109553	1.229241	0.222
C(7)	1.155596	0.855702	1.350465	0.180
R-squared	0.689825	Mean dep var		0.24323
Adjusted R-squared	0.668185	S.D. dep var		0.52727
S.E. of regression	0.303729	Akaike IC		0.52692
Sum squared resid	7.933607	Schwarz IC		0.71754
Log likelihood	-17.50192	Hannan-Quinn		0.60389
F-statistic	31.87716	DW stat		2.12359
Prob(F-statistic)	0.000000			

Table 26: *Estimation results for the AR(7) model for $rcpi$ in the estimation period containing the global financial crisis.*

The Crisis Period

- This specification yields an adjusted $R^2 = 0.67$, substantially higher than the one for the pre-crisis period (which was 0.38), with a Durbin Watson statistic of 2.12.
- Figures 30 and 31, and Table 27 report the diagnostic tests on the residuals that overall support the model specification.

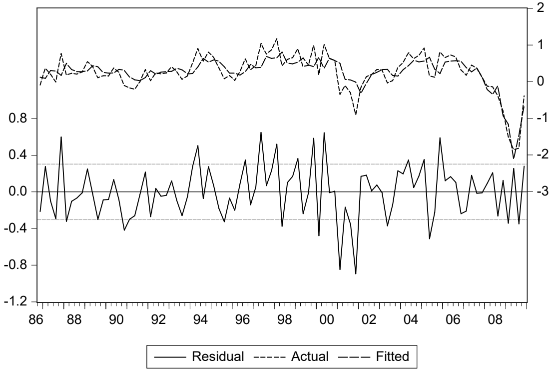
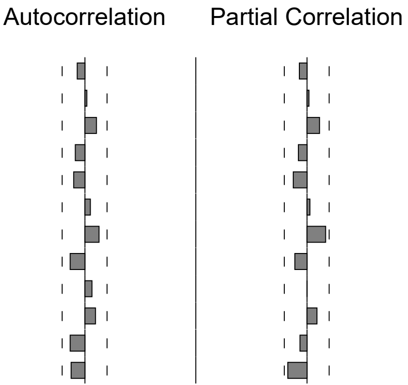


Figure 30: *The residuals of the estimated AR(7) model for* $rcpi$ *in the extended sample including the crisis*



	AC	PAC
1	-0.070	-0.070
2	0.024	0.019
3	0.112	0.115
4	-0.089	-0.075
5	-0.106	-0.126
6	0.053	0.031
7	0.134	0.174
8	-0.138	-0.110
9	0.066	0.004
10	0.105	0.093
11	-0.132	-0.064
12	-0.130	-0.173

Figure 31: *Correlogram of the residuals of the AR(7) model for* $rcpi$

Diagnostic checking on the residuals from AR(7)

Breusch-Godfrey			
Serial Correlation LM Test:			
F-statistic	3.429180	F(2,84)	0.037
Obs*R-squared	7.020020	Chi-Sq(2)	0.029
Heteroskedasticity Test: ARCH			
F-statistic	0.025863	F(1,90)	0.872
Obs*R-squared	0.026431	Chi-Sq(1)	0.870
Heteroskedasticity Test: White			
F-statistic	0.188135	F(14,78)	0.999
Obs*R-squared	3.037834	Chi-Sq(14)	0.999
Scaled explained SS	3.384554	Chi-Sq(14)	0.998

Table 27: *Diagnostic checking on the residuals from AR(7) model*

The Crisis Period

It is interesting to note how the dummy variable affects the estimated parameter values, since it is as if we had two separate equations

$$\text{Before the crisis:} \quad rcpi_t = 0.14 + 0.46 rcpi_{t-1} + 0.13 rcpi_{t-2} + \varepsilon_{1t} \tag{47a}$$

$$\begin{aligned} \text{During the crisis:} \quad rcpi_t = & -0.43 + 2.88 rcpi_{t-1} - 2.76 rcpi_{t-2} \\ & + 1.15 rcpi_{t-7} + \varepsilon_{2t} \end{aligned} \tag{47b}$$

- In particular, the persistence (as measured by the sum of the AR coefficients) increases substantially during the crisis period.
- Computing forecasts with this specification amounts to exploiting the information of both models (47a) and (47b) combined into a single specification.

Forecasts from AR(7) model for $rcpi$ against the actuals 2010 - 2012

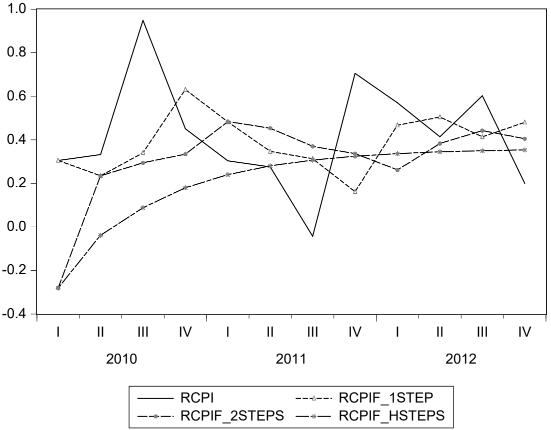


Figure 32: *Forecasts from AR(7) model for $rcpi$ against the actuals 2010 - 2012*

- The overall impression from these results is that the forecast performance remains overall stable after the crisis, once the model is appropriately modified to allow for parameter changes during the crisis.
- It is interesting to consider what would happen if we did not include the dummies for the financial crisis, and just produced forecasts with a simple AR(2) model.
- The estimation of the model without augmentation for the dummy produces the following equation:

$$rcpi_t = 0.048 + 0.7rcpi_{t-1} + 0.07rcpi_{t-2} + \varepsilon_{1t}. \tag{48}$$

- The associated residuals, graphed in Figure 33, are large and negative during the global financial crisis.
- The proceeding slides show the forecasts obtained through model (48) against those we obtained through the richer specification (47a)-(47b), both at one-, two-, and one- to h-steps ahead.

Residuals of the mis-specified model

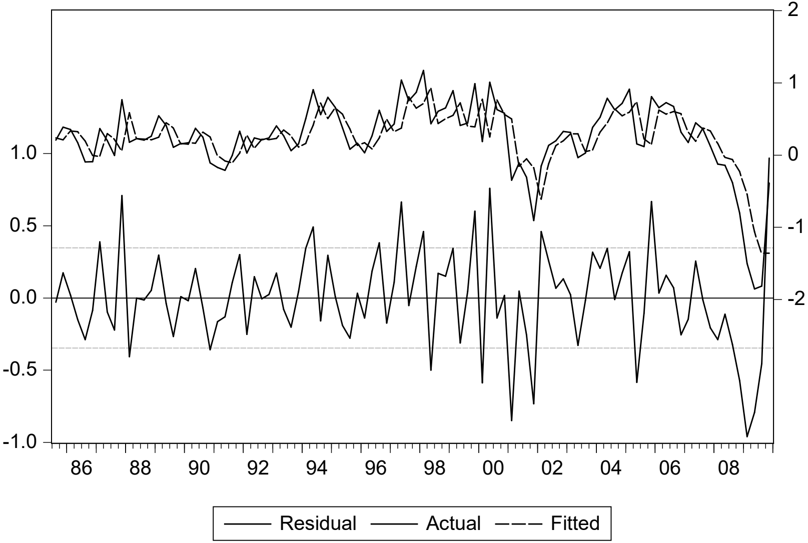


Figure 33: *Residuals of the mis-specified model which does not contain the dummy variable*

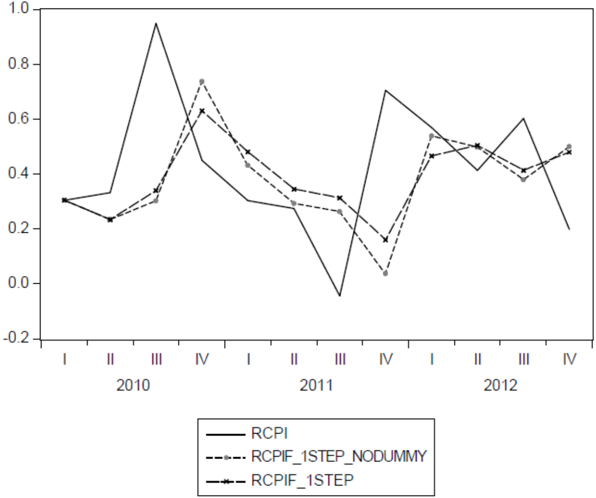


Figure 34: *1-step forecasts not accounting for the crisis (no dummy) against those obtained through the richer specification containing the dummy, against the actuals*

h-steps

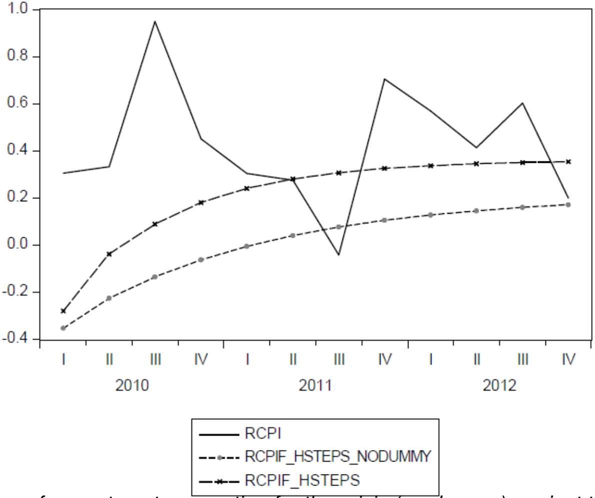


Figure 35: *h-steps forecasts not accounting for the crisis (no dummy) against those obtained through the richer specification containing the dummy, against the actuals*

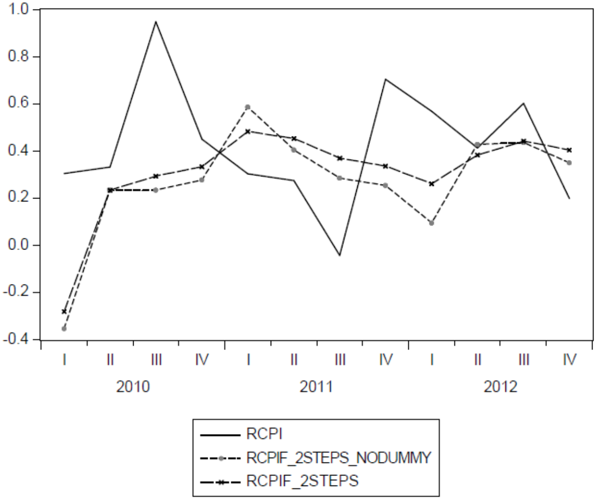


Figure 36: *2-steps forecasts not accounting for the crisis (no dummy) against those obtained through the richer specification containing the dummy, against the actuals*

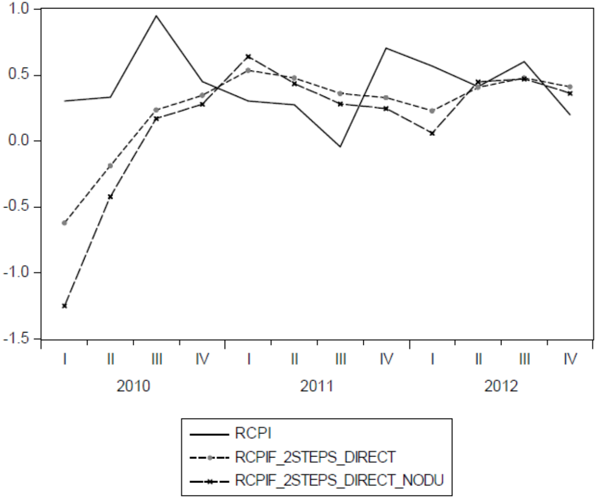


Figure 37: *2-steps direct forecasts not accounting for the crisis (no dummy) against those obtained through the richer specification containing the dummy, against the actuals*

Concluding Remarks

The models we have considered are overall rather simple, since they basically explain the variable of interest using its past only. Hence, a natural question is whether these univariate time series methods are useful in practice. The answer is yes, for the following reasons.

- From a theoretical perspective, any weakly stationary process (or integrated process after proper differencing) can be written as an MA(∞), and under mild conditions the latter can be approximated by an ARMA model.
- The high persistence of several economic variables suggests that forecasting methods that exploit the present and the past behavior of the variable to predict the future can perform well.

Concluding Remarks

- Forecasts from these models can provide a benchmark for comparison of more elaborate forecasts or can be combined with forecasts from other models to assess whether a lower loss function can be obtained.
- Forecast failure of these models can provide an indication of the type of information that is missing in more elaborate models, e.g., it can suggest that a substantial amount of dynamics should be included.
- More practically, empirically these forecasting methods tend to perform well for a variety of variables and across different evaluation samples, when used for short-term forecasting.