# MFFM ECON5022

## Lecture 2: Linear Processes

Dr. Minjoo Kim

Economics
University of Glasgow

# White noise

- A time series (stochastic process) $\epsilon_t$ is called a white noise (WN) process if it is a sequence of independent and identically distributed random variables with zero mean and finite variance.
- We will usually assume Gaussian WN processes, which we write in the following form

$$\epsilon_t \sim iidN(0, \sigma^2) \tag{1}$$

with the following properties

1. $E\left[\epsilon_t\right] = E\left[\epsilon_t | \epsilon_{t-1}, \ldots, \epsilon_{t-p}\right] = 0$
2. $Var\left(\epsilon_t\right) = Var\left(\epsilon_t | \epsilon_{t-1}, \ldots, \epsilon_{t-p}\right) = \sigma^2$
3. $E\left[\epsilon_t, \epsilon_s\right] = Cov\left(\epsilon_t, \epsilon_s\right) = 0$ for all $t \neq s$

- These properties imply the absence of serial correlation or predictability (1-3), and conditional homoskedasticity (2).

# Conditional and unconditional quantities

- The unconditional mean and variance of any time series $y_t$ are $E[y_t]$ and $Var(y_t)$, respectively, i.e. NOT conditional on the past of the series.
- We can also estimate the conditional mean and variance. These are conditional to some time-series model assumed, i.e. conditional to the past values.
- We can denote these as

$$E[y_t|y_{t-1},\ldots,y_1] \text{ and } Var(y_t|y_{t-1},\ldots,y_1)$$

or as

$$E[y_t|\mathcal{F}_{t-1}] \text{ and } Var(y_t|\mathcal{F}_{t-1})$$

where $\mathcal{F}_t$ is the "information set", i.e. everything we know up to time $t$ (i.e. the past lags of $y_t$).

# A few comments about the WN model

- We can easily see that the WN model, $y_t = \epsilon_t$, is not a proper model for our time series $y_t$.
- The unconditional mean (variance) is equal to the conditional mean (variance).
- The unconditional mean (variance) can simply be estimated from the sample quantities (i.e. sample mean and variance).
- There is no correlation of $\epsilon_t$ with future ($\epsilon_{t+k}$) or past ($\epsilon_{t-k}$) values, hence the WN is not good to model dependence with the past, and to forecast the future.

# Autoregressive (AR) process

- The first step to model dependence in the mean of a time series is to use models where the mean of the time series $y_t$ depends on past information, i.e. to consider models of the form

$$y_t = E\left[y_t | \mathcal{F}_t\right] + \epsilon_t. \tag{2}$$

- The simplest form of dependence of $y_t$ on $\mathcal{F}_t := \{y_{t-1}, y_{t-2}, \ldots, y_2, y_1\}$, is of the linear form. Hence we can define the model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t, \tag{3}$$

which is called the "autoregressive model of order $p$" or AR(p) for short, since $y_t$ is regressed on itself/its past values (auto-regression).

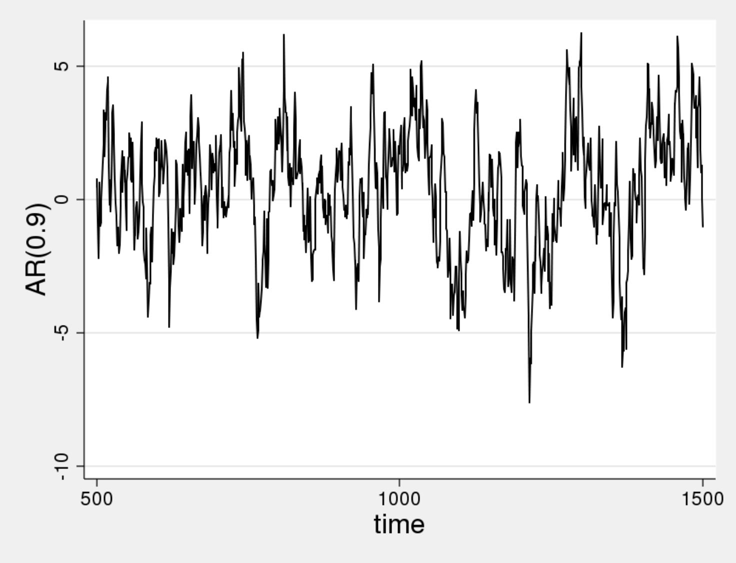



Figure: AR $\phi = 0.9$

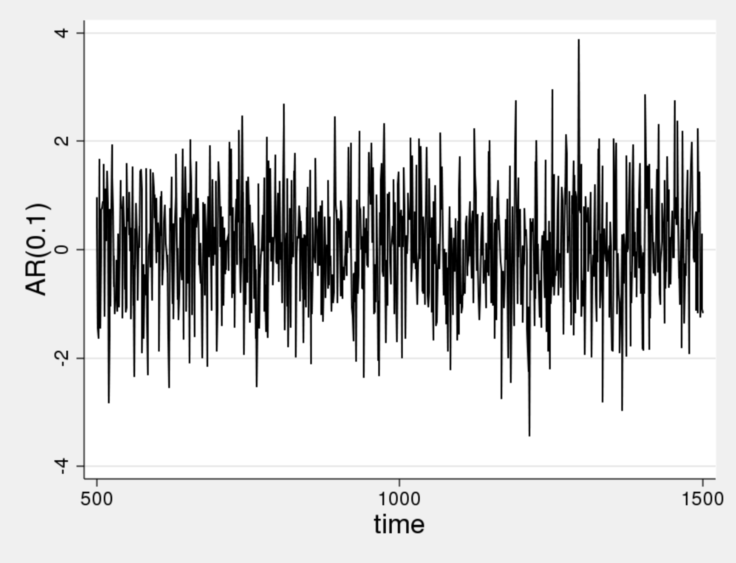



Figure: AR $\phi = 0.1$

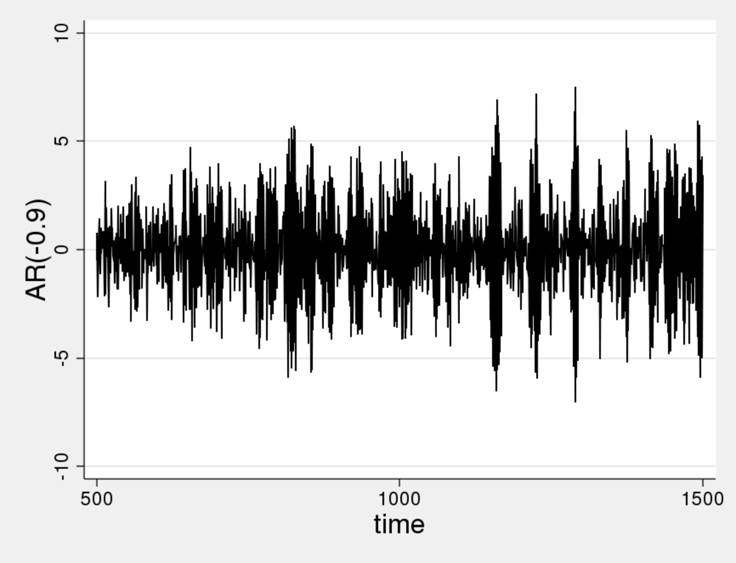



Figure: AR $\phi = -0.9$

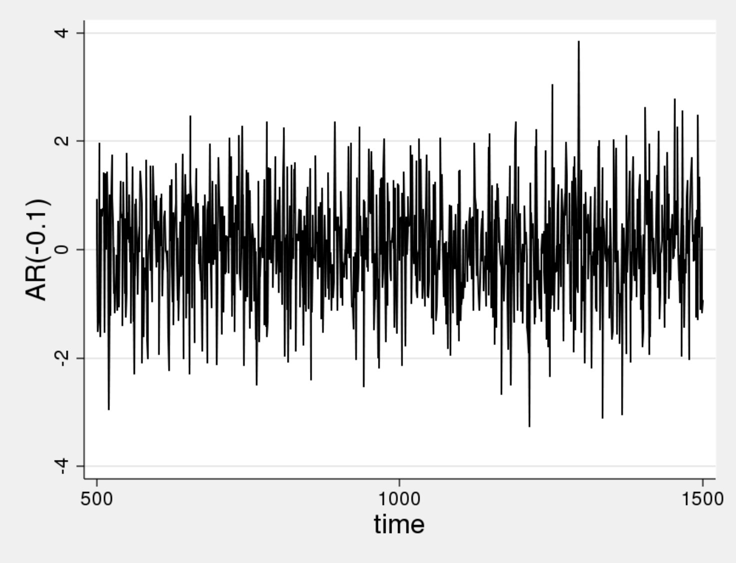



Figure: AR $\phi = -0.1$

# Properties of AR process

## Unconditional mean

- Consider the case of the AR(1) model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t. \tag{4}$$

- The conditional mean is

$$E[y_t|y_{t-1}] = \phi_0 + \phi_1 E[y_{t-1}|y_{t-1}] = \phi_0 + \phi_1 y_{t-1}. \tag{5}$$

- The unconditional mean is

$$\begin{aligned} E[y_t] &= \phi_0 + \phi_1 E[y_{t-1}] + E[\epsilon_t] \\ &= \phi_0 + \phi_1 E[y_t] + 0 \\ \Rightarrow E[y_t] &= \frac{\phi_0}{1-\phi_1}. \end{aligned} \tag{6}$$

- For a stationary process it holds that

$$E[y_t] = E[y_{t-1}] \tag{7}$$

i.e. The mean remains the same for different subsamples.

- This shows why it is important to be able to work with stationary time series: many of the models we are using have nice properties that hold only for stationary time series.
- For a nonstationary time series,

$$E[y_t] \neq E[y_{t-1}] \tag{8}$$

and hence the unconditional mean does not exist!

# Properties of AR process

## Unconditional variance

- The conditional variance of the AR(1) model is

$$Var(y_t|\mathcal{F}_t) = \sigma^2. \tag{9}$$

- The unconditional variance of the AR(1) model is

$$\begin{aligned} Var(y_t) &= Var(\phi_0) + \phi_1^2 Var(y_{t-1}) + Var(\epsilon_t) \\ &= 0 + \phi_1^2 Var(y_t) + \sigma^2 \\ &\Rightarrow Var(y_t) = \frac{\sigma^2}{1-\phi_1^2} \end{aligned} \tag{10}$$

where $Var(y_t) = Var(y_{t-1})$ for stationary processes!

# Properties of AR processes

## Repeated substitution

- We can solve the AR(1) model using repeated substitution to obtain:

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \epsilon_t \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &\vdots \\ &= \phi_0(1 + \phi_1 + \phi_1^2 + \cdots) + \phi_1^\infty y_0 + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \\ &= \phi_0(1 - \phi_1)^{-1} + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \end{aligned} \tag{11}$$

since for $|\phi_1| < 1$ it holds that:

$$\begin{gathered} 1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^\infty \cong 1/(1 - \phi_1) \\ \phi_1^\infty \to 0. \end{gathered} \tag{12}$$

- We assumed that the process is observed for an infinite amount of time, but we can have similar results if the sample is $T$ but it is large.
- This is a neat way to write any AR(1) process for stationary time series, as a function of past shocks $(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$.
- The restriction we put for solution is that $|\phi_1| < 1$. We will see next time that this is the restriction for the AR(1) process to be stationary.

- Find the unconditional mean and variance of the AR(1) process using the formula in the previous page.
  - The formula from repeated substitution is

$$y_t = \phi_0(1-\phi_1)^{-1} + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}.$$

  - Then the uncoditional mean is:

$$E[y_t] = E\left[\phi_0(1-\phi_1)^{-1}\right] + \sum_{i=0}^{\infty} \phi_1^i E[\epsilon_{t-i}] = \frac{\phi_0}{1-\phi_1}$$

  - And the unconditional variance is:

$$Var(y_t) = \sum_{i=0}^{\infty} \phi_1^{2i} Var(\epsilon_{t-i}) = \frac{\sigma^2}{1-\phi_1^2}$$

- Using the formula derived with repeated substitution, show that the $k$-th order autocovariance function of an AR(1) model is

$$Cov(y_t, y_{t-k}) = E[y_t y_{t-k}] = \left(\frac{\sigma^2}{1-\phi_1^2}\right)\phi_1^k \tag{13}$$

under the assumption of $E[y_t] = E[y_{t-k}] = 0$.

- Using the formula for the autocovariance and variance, show that the $k$-th order autocorrelation of an AR(1) is given by

$$Corr(y_t, y_{t-k}) = \phi_1^k. \tag{14}$$

- An **autoregressive process** has
  1. A geometrically decaying ACF,
  2. Number of spikes of PACF = AR order.

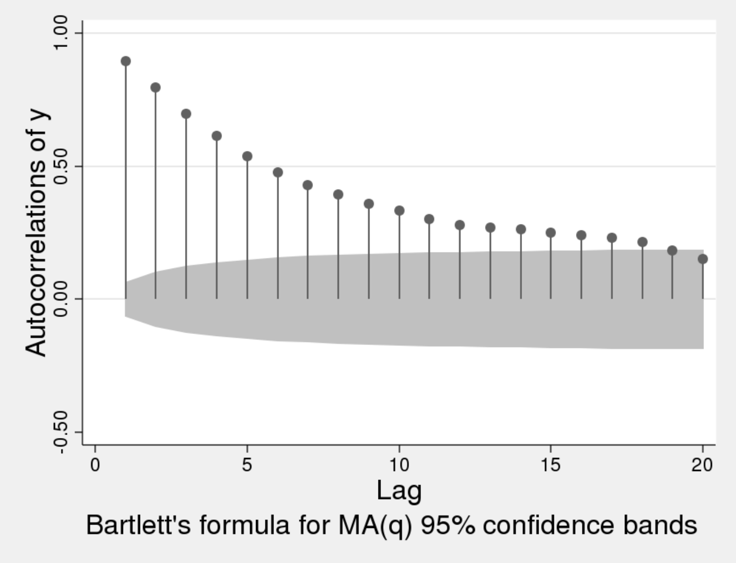



Figure: AR $\phi = 0.9$

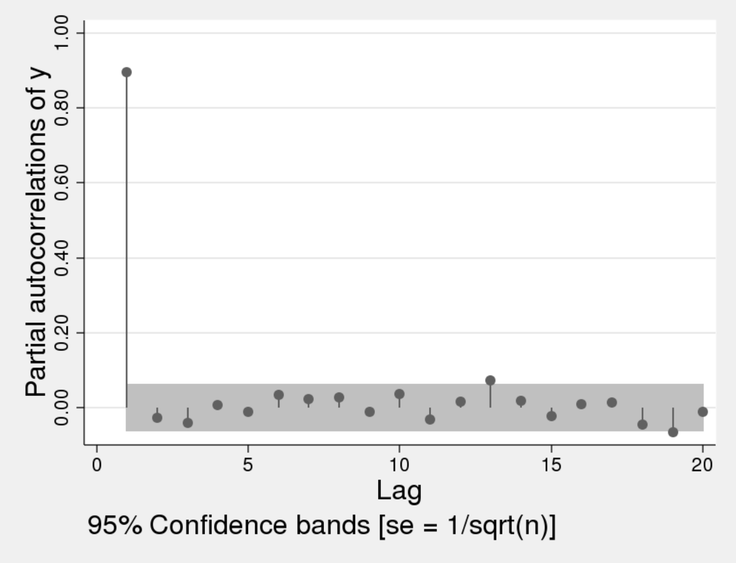



Figure: AR $\phi = 0.9$

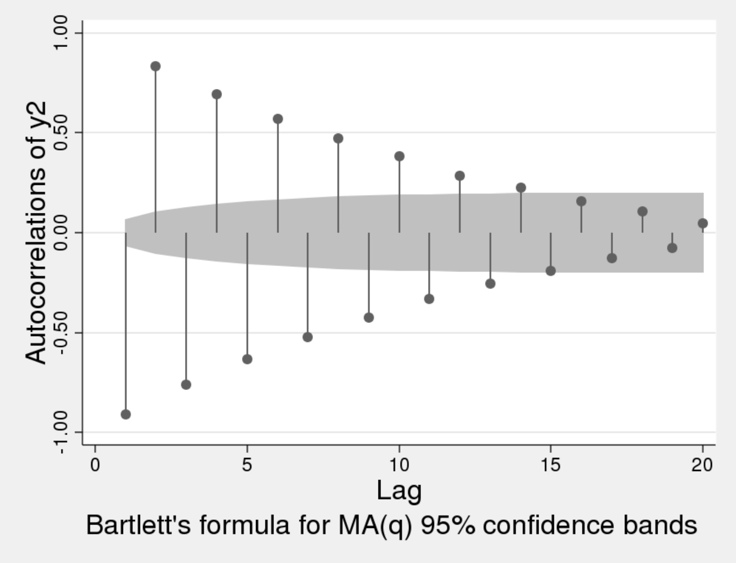



Figure: AR $\phi = -0.9$

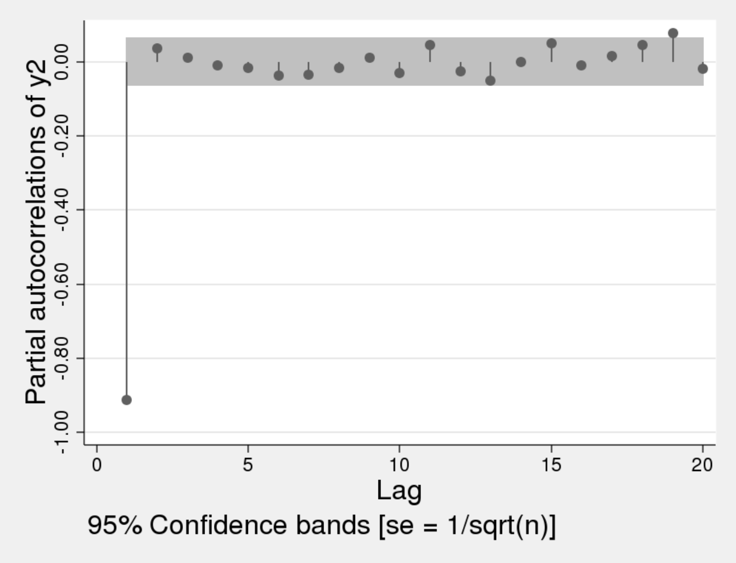



Figure: AR $\phi = -0.9$

# Moving Average Processes

- Let $\epsilon_t$ be a sequence of independently and identically distributed (iid) random variables with $E[\epsilon_t] = 0$ and $Var(\epsilon_t) = \sigma^2$, then

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$$

  is a $q$th order moving average model MA(q).

- Its properties are

$$\begin{aligned} E[y_t] &= \mu \\ Var(y_t) &= (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2 \\ Cov(y_t, y_{t-s}) &= \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \cdots + \theta_q\theta_{q-s})\sigma^2, & \text{for } s \leq q \\ 0, & \text{for } s > q \end{cases} \end{aligned}$$

- A **moving average process** has
  1. Number of spikes of ACF = MA order,
  2. A geometrically decaying PACF.

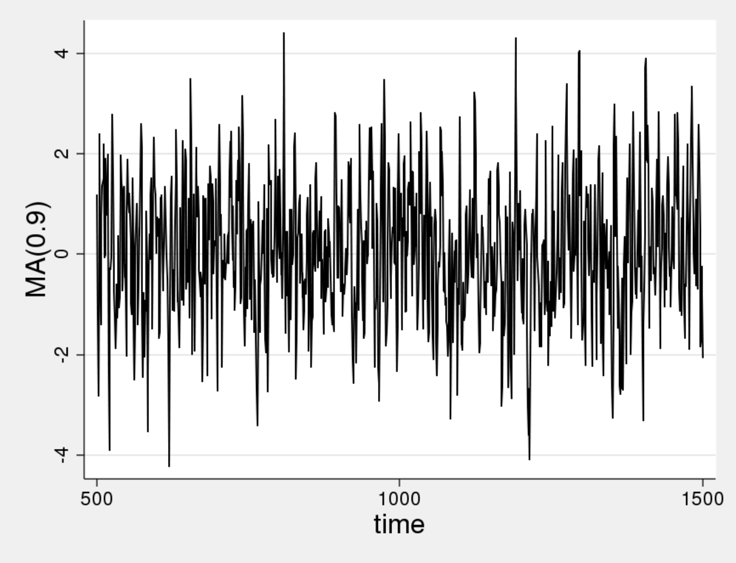



Figure: MA $\theta = 0.9$

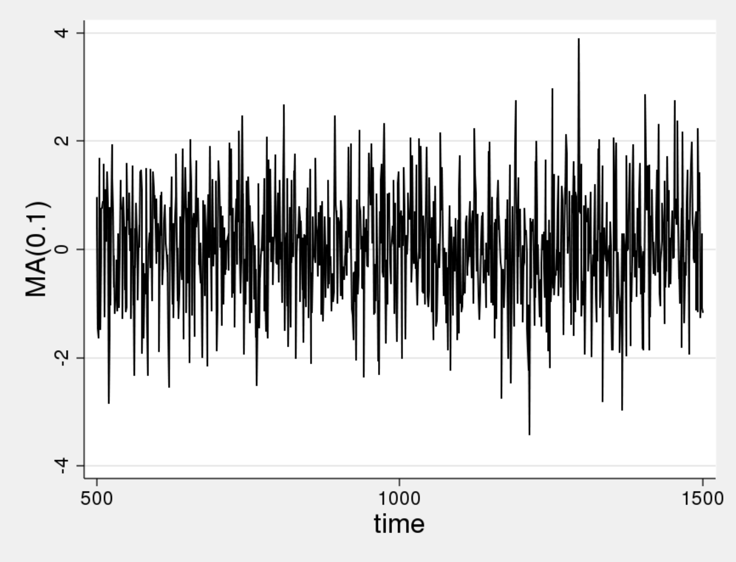



Figure: MA $\theta = 0.1$

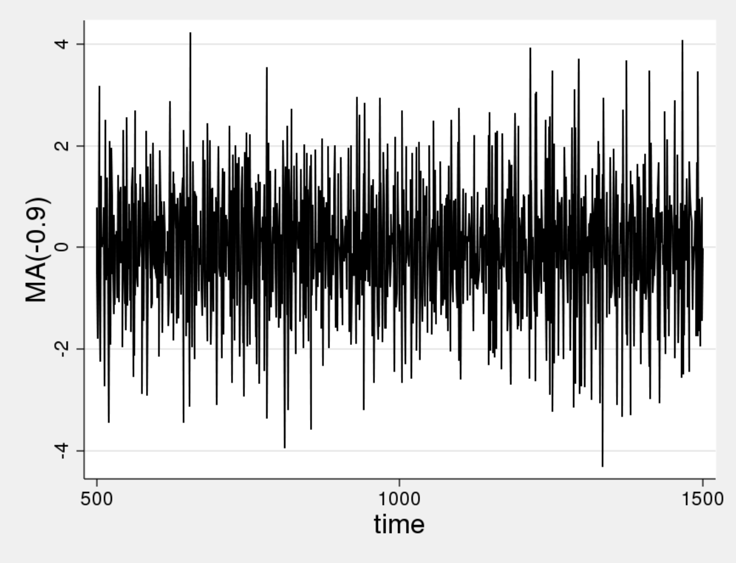



Figure: MA $\theta = -0.9$

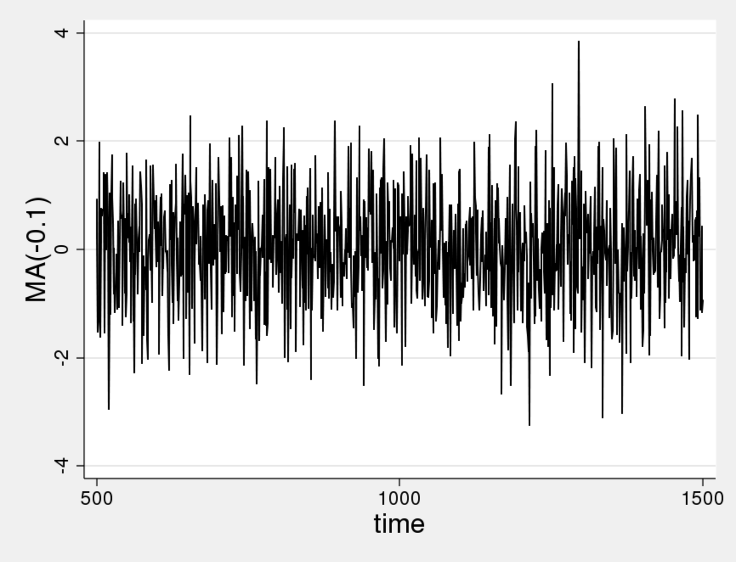



Figure: MA $\theta = -0.1$

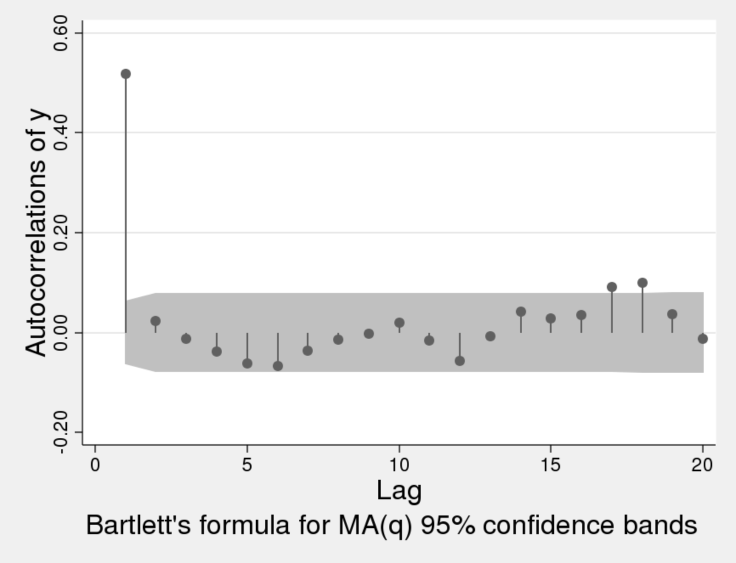



Figure: MA $\theta = 0.9$

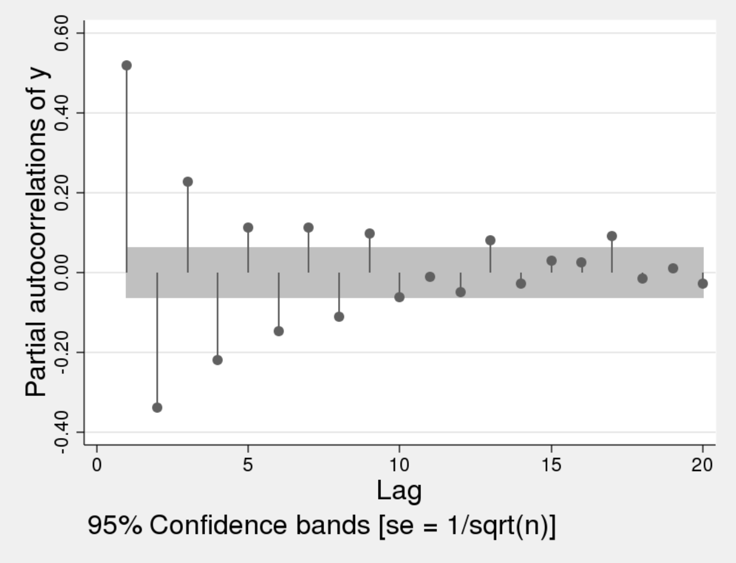



Figure: MA $\theta = 0.1$

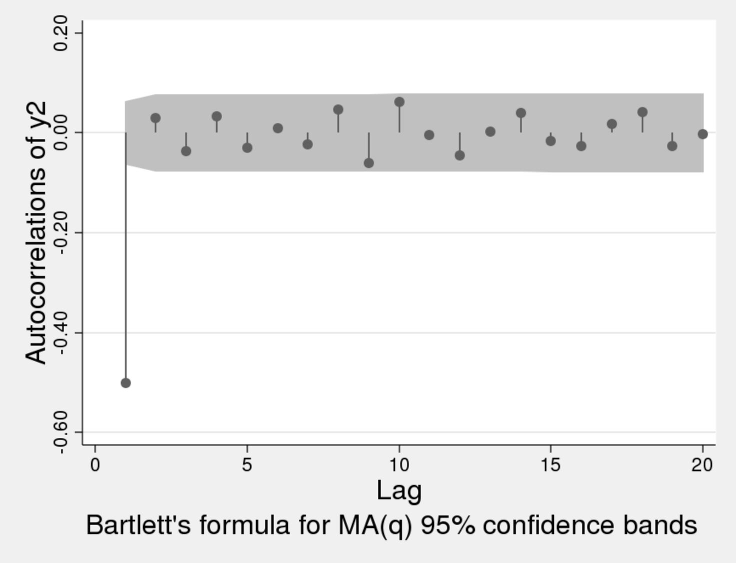



Figure: MA $\theta = -0.9$

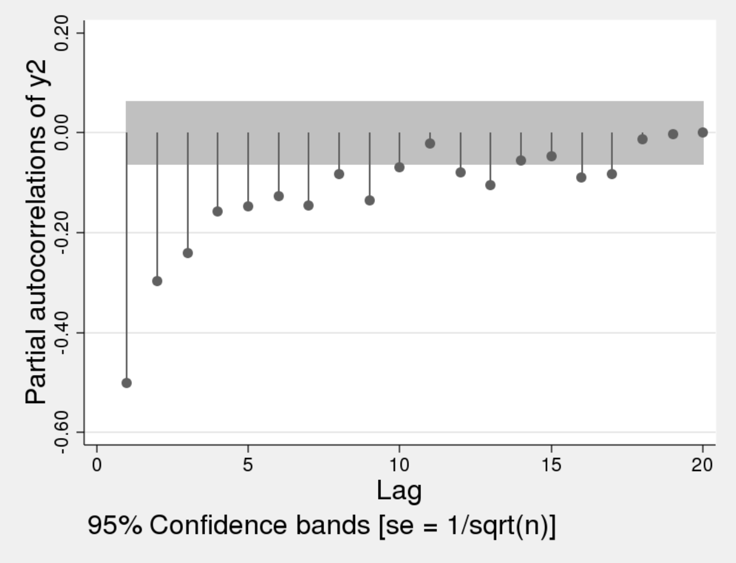



Figure: MA $\theta = -0.1$

# Example of an MA problem

Consider the following MA(2) process:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \tag{15}$$

where $\epsilon_t$ is a zero mean white noise process with variance $\sigma^2$.

1. Calculate the mean and variance of $y_t$.
2. Derive the autocorrelation function for this process (i.e. express the autocorrelations, $\rho_1, \rho_2, \ldots$ as functions of the parameters $\theta_1$ and $\theta_2$).
3. If $\theta_1 = -0.5$ and $\theta_2 = 0.25$, sketch the autocorrelation function (ACF) of $y_t$.

# ARMA process

- AR(p) and MA(q) processes alone can become cumbersome, because a higher order lags $p$ and $q$ might be needed. (Why? See Exercise 3).
- We can combine both processes into one, called Autoregressive Moving-Average, or ARMA(p,q) for short.
- The simplest case, the ARMA(1,1) process takes the form

$$y_t - \phi_1 y_{t-1} = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{16}$$

- The crucial condition for this model to make sense is $\phi_1 \neq -\theta_1$ (Why? See Exercise 2)

# Properties of ARMA processes

- For the simple ARMA(1,1) process (we can then easily generalize to ARMA(p,q) processes) we have

$$E[y_t] = \frac{\mu}{1-\phi_1} \tag{17}$$

$$Var(y_t) = \frac{(1+2\phi_1\theta_1+\theta_1^2)\sigma^2}{1-\phi_1^2} \tag{18}$$

$$Cov(y_t, y_{t+k}) = \phi_1^k Var(y_t) + \phi_1^{k-1}\theta_1\sigma^2, \text{ for } k \geq 1 \tag{19}$$

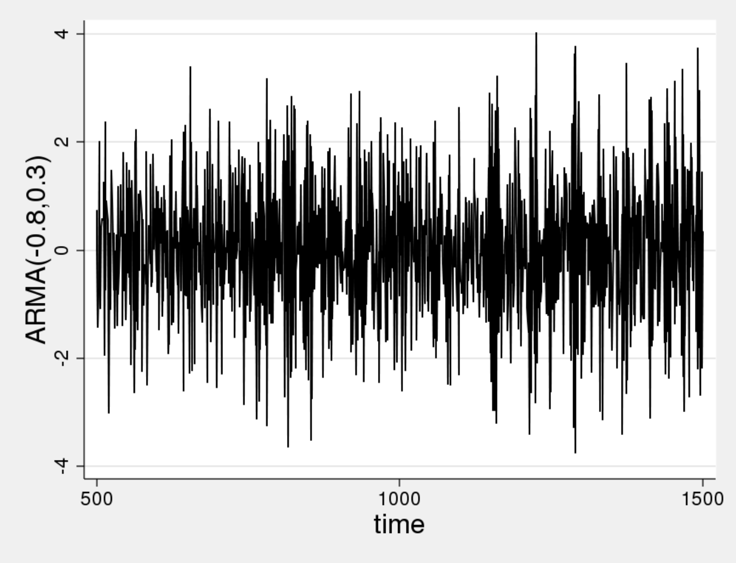



Figure: ARMA $\phi = -0.8, \theta = 0.3$

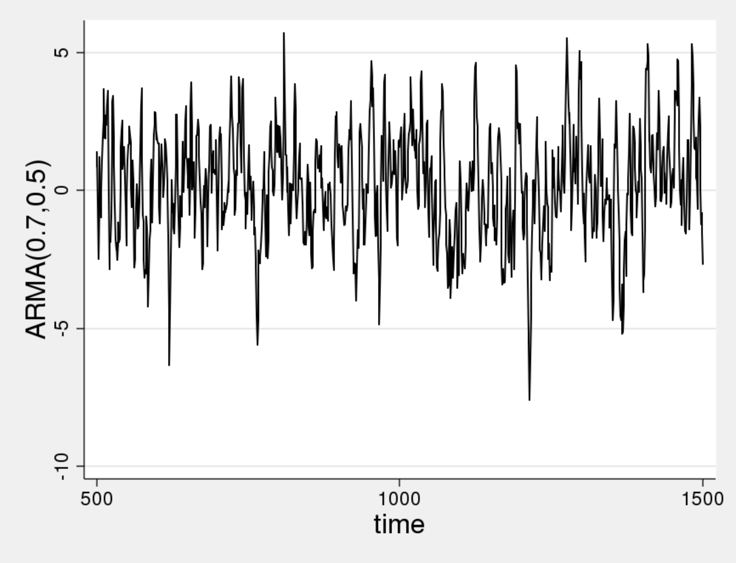



Figure: ARMA $\phi = 0.7, \theta = 0.5$

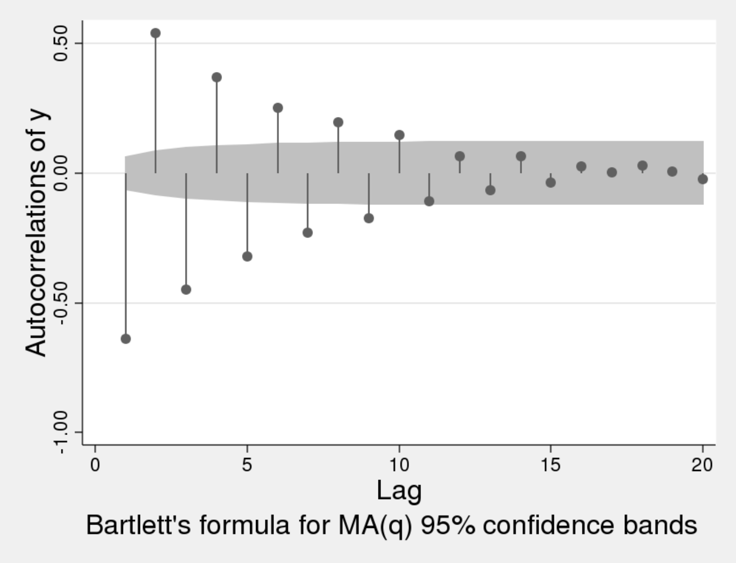



Figure: ARMA $\phi = -0.8, \theta = 0.3$

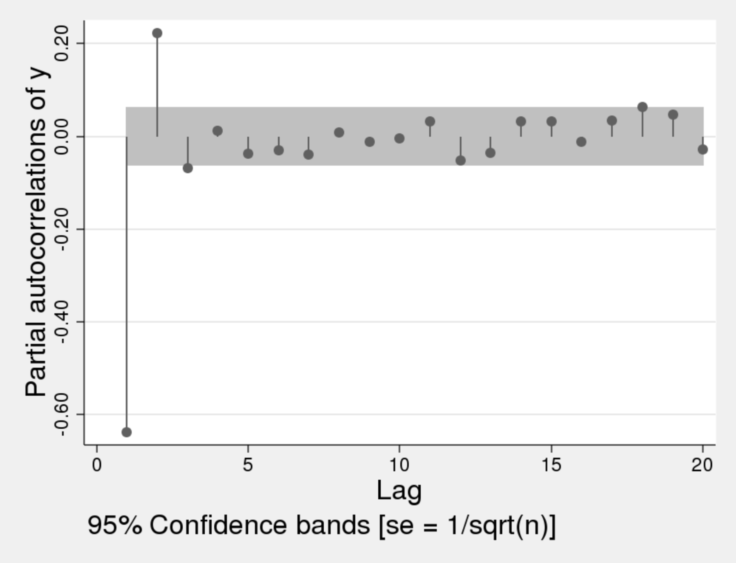



Figure: ARMA $\phi = -0.8, \theta = 0.3$

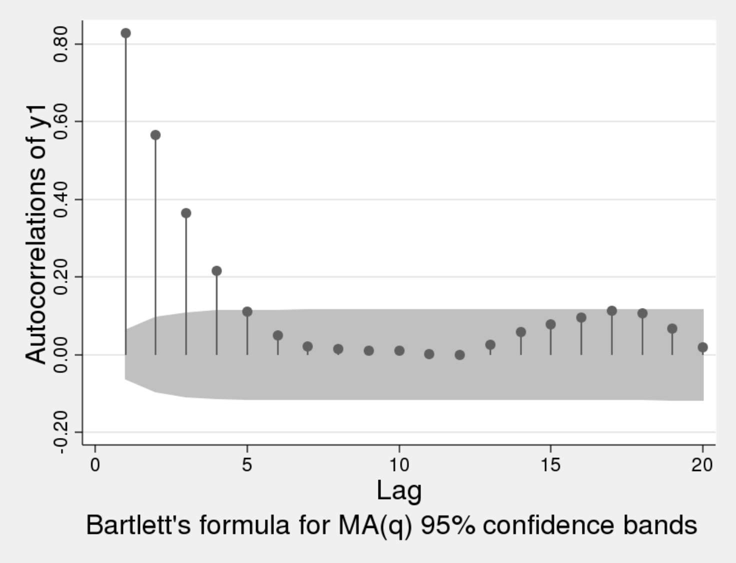



Figure: ARMA $\phi = 0.7, \theta = 0.5$

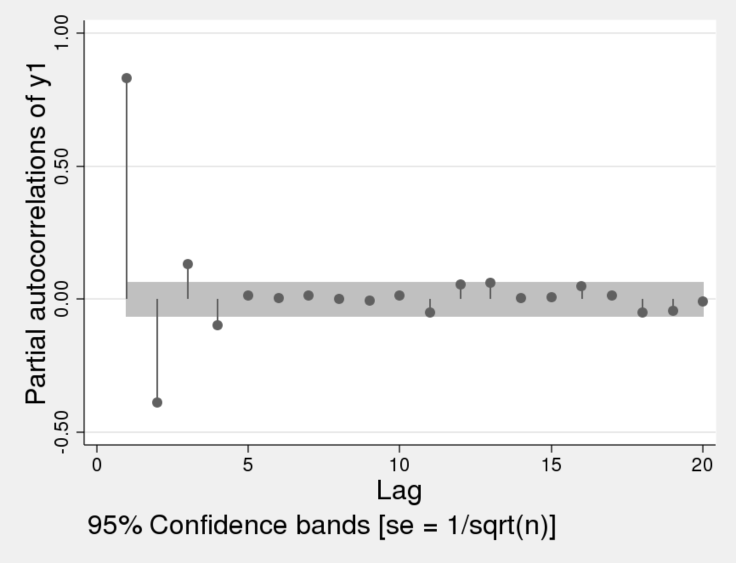



Figure: ARMA $\phi = 0.7, \theta = 0.5$

# Stationarity of an AR process

- The AR(p) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{20}$$

is stationary, if the roots of the characteristic polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \tag{21}$$

are all larger than 1 in absolute value.

- A stationary AR(p) model has an MA($\infty$) representation (**Invertibility**). For example, in the AR(1) model,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t,$$

characteristic polynomial is

$$1 - \phi_1 z = 0 \Rightarrow z = 1/\phi_1$$
$$|z| > 1 \Rightarrow |\phi_1| < 1.$$

# Random walk model

- The random walk model

$$y_t = y_{t-1} + \epsilon_t \tag{22}$$

is a special case of the AR(1) model. From the previous slide we can see this model is non-stationary since $\phi_1 = 1$ (ACF?).

- From repeated substitution we obtain

$$y_t = y_0 + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots \tag{23}$$

The past shocks $e_{t-s}$ for any $s > 0$ are just added to the current (time $t$) value and they don't "DIE OUT".

- In contrast, in the AR(1) model we had

$$y_t = y_0 + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots \tag{24}$$

For more distant shocks $\phi_1^s \to 0$ as $s \to \infty$, since $|\phi_1| < 1$.

# General processes with unit root ($|\phi_1| = 1$)

- The random walk case is quite interesting empirically (we will see this with real data), despite the fact that it describes a non-stationary process. The case where $|\phi_1| > 1$ is not interesting, since this process is highly explosive and we cannot do anything with it (unless we transform it to stationarity).
- Obviously we can generalize to the case of more unit roots, for instance the model

$$y_t = y_{t-1} - y_{t-2} + \epsilon_t \tag{25}$$

  has two unit roots but this is hardly a case you will meet (empirically).
- Testing for a unit root in the coefficient of $y_{t-1}$ (one unit root) is done with the (Augmented) Dickey-Fuller test or others.

# Estimation of ARMA processes

- Estimation of AR, MA and ARMA models can be done with OLS or maximum likelihood or Bayesian methods, depending on the model.
- In this course we will not cover estimation methods; modern econometric software (OxMetrics, Eviews, Stata) does estimation for us.
- Each model has a likelihood function, for instance the AR(1) model can be written as

$$y_t | y_{t-1} \sim N(\phi_0 + \phi_1 y_{t-1}, \sigma^2)$$

which implies the likelihood function

$$\mathcal{L}(y|\phi_0, \phi_1, \sigma^2) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_t - \phi_0 - \phi_1 y_t}{\sigma}\right)^2}$$

# Fitting ARMA processes

- The likelihood function shows "how likelily" is a specific combination of parameters $\phi_0, \phi_1, \sigma^2$ to have generated the process $y_t$. The larger the value of the likelihood function, the better.
- Adding more lags, exogenous explanatory variables, trend and dummy variables, increases the likelihood. More correlated (important) variables increase the value of the likelihood a lot, unimportant variables increase the value of the likelihood a little bit (or not at all if they have exactly zero correlation with $y_t$). However, we usually have finite samples, so adding many lags can reduce the degrees of freedom. For instance, if we fit an AR(1) we lose 1 observation, if we fit an AR(100) we lose 100 observations.

- Additionally, even if we have 100,000 observations, fitting an AR(100) can fit the data very well in sample, but do very bad out-of-sample (i.e. when forecasting).
- Example: Assume you have a very flexible model with many lags that fits your data very well for the period 1990-2007 and which you use to forecast 2008. Once the 2008 crisis has been observed ("which is a rare event"), you most probably have made a very bad forecast using this flexible model.
- This problem is known as "over-fitting". In general, given a sample of $T$ observations: Among two models with equal performance we always should chose the most parsimonious one (the one with less coefficients to estimate).

# Model selection

- Hence we need to select a model that fits the data well (high value of the likelihood), but also is small in size.
- *Information Criteria* do exactly this. They are the sum of the value of the likelihood at the estimated parameters, and a term which penalizes complex models.
- Depending on the exact functional form of the penalty, we can define several criteria:

$$
\begin{aligned}
AIC &= -2\ln(\mathcal{L}) + 2k \\
SBIC &= -2\ln(\mathcal{L}) + k\ln(T) \\
HQIC &= -2\ln(\mathcal{L}) + 2k\ln(\ln(T))
\end{aligned}
$$

where $k$ denotes the number of parameters and $T$ the sample size available to estimate these parameters.

# Forecasting

- Forecasting $=$ prediction.
- An important test of the adequacy of a model. e.g.
  - Forecasting tomorrow's return on a particular share
  - Forecasting the price of a house given its characteristics
  - Forecasting the riskiness of a portfolio over the next year
  - Forecasting the volatility of bond returns
- We can distinguish two approaches:
  - Econometric (structural) forecasting
  - Time series forecasting

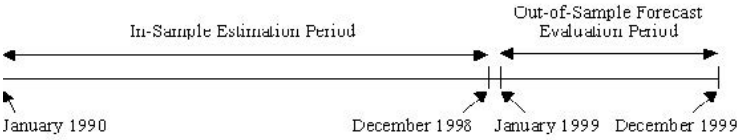
  The distinction between the two types is somewhat blurred (e.g., VARs).

Why use forecasting methods to evaluate models?

1. Less prone to data mining, in-sample over-fitting (indeed, in-sample over-fitting leads to out-of-sample "under-fitting")
2. Instability (the coefficients estimated in-sample might not be optimal if a structural break, e.g. financial crisis, occurs during the forecast period)
3. In-sample methods too difficult to implement/trust for very complete models
4. This is the end purpose of most time-series modelling, hence, "all models are false" but at the same time a "model is as good as its forecasts".

# In-sample vs. out-of-sample (OOS)

Say we have some data - e.g. monthly FTSE returns for 120 months: 1990M11 - 1999M12. We could use all of it to build the model, or keep some observations back:




A good test of the model since we have not used the information from 1999M1 onwards when we estimated the model parameters.

# Pseudo out-of-sample (POOS)

## Recursive forecasting

For simplicity, assume AR(1): $y_t = \phi y_{t-1} + \epsilon_t$. Sample size is $T = R + P$. Last $P$ periods used for 'prediction' (construction of pseudo-out-of-sample forecasts). Strategy:

1. Estimate $\phi$ using observations $1 : R$, $\hat{\phi}_R$
2. Forecast $y_{R+1}$ using $y_R$ and $\hat{\phi}_R$.
3. Estimate $\phi$ using observations $1 : R+1$, $\hat{\phi}_{R+1}$
4. Forecast $y_{R+2}$ using $y_{R+1}$ and $\hat{\phi}_{R+1}$.
5. Estimate $\phi$ using observations $1 : R+2$, $\hat{\phi}_{R+2}$
6. Forecast $y_{R+3}$ using $y_{R+2}$ and $\hat{\phi}_{R+3}$.
7. Do this until you "exhaust" the sample of $P$ observations.

# Pseudo out-of-sample (POOS)

## Rolling forecasting

Same model, same sample as previous slide. Assume you are forecasting using the most recent 60 observations (i.e. a "rolling window" of 60 observations). Strategy:

1. Estimate $\phi$ using observations $R-60:R$, $\hat{\phi}_R$
2. Forecast $y_{R+1}$ using $y_R$ and $\hat{\phi}_R$.
3. Estimate $\phi$ using observations $R-59:R+1$, $\hat{\phi}_{R+1}$
4. Forecast $y_{R+2}$ using $y_{R+1}$ and $\hat{\phi}_{R+1}$.
5. Estimate $\phi$ using observations $R-58:R+2$, $\hat{\phi}_{R+2}$
6. Forecast $y_{R+3}$ using $y_{R+2}$ and $\hat{\phi}_{R+3}$.
7. Do this until you "exhaust" the sample of $P$ observations.

# Forecasting basics

- $y_{t+h}$: variable to be forecast
- $x_t$: vector of used to make forecast (typically would include current and lags of $y_t$ and other variables)
- $f_{t+h|t}$: forecast of $y_{t+h}$ made at time $t$
- $e_{t+h} = y_{t+h} - f_{t+h|t}$: forecast error
- $L(e_{t+h})$: loss associated with the forecast error
- $E[L(e_{t+h})]$: "risk" associated with the forecast

# Iterated vs direct forecasts

Example: AR(1) model, $y_t = \phi y_{t-1} + \epsilon_t$

- Goal: forecast $y_{t+2}$
- Optimal forecast: $f_{t+2|t} = \beta y_t$, where $\beta = \phi^2$

**Two ways to forecast $y_{t+2}$**

- "Iterated": Estimate $\phi$ from $y_t = \phi y_{t-1} + \epsilon_t$, and use $f_{t+2|t}^{iterated} = \hat{\phi}^2 y_t$.
- "Direct": Estimate $\beta$ from $y_t = \beta y_{t-2} + u_t$, and use $f_{t+2|t}^{direct} = \hat{\beta} y_t$.

Pros and Cons:

- $\hat{\phi}$ is MLE, and is "efficient" under $AR(1)$, but what if model is misspecified?
- $\hat{\beta}$ has larger variance than $\hat{\phi}$ under correct specification, but is robust to misspecification (for the class of forecasts under consideration).

Literature: Cox (1961) to Schorfheide (2005) (survey: Bhansali, 1999)

Empirical comparison: Marcelino, Stock and Watson (2006)

- 170 monthly U.S. macro series, 1959-2002
- Pseudo-out-of-sample forecasts (POOS)

# Backshift/lag operator

- The properties of the lag operator $L$ (sometimes denoted as backshift operator, $B$) are:

$$\begin{aligned} Lx_t &= x_{t-1} \\ L^2 x_t &= L(Lx_t) = Lx_{t-1} = x_{t-2} \\ &\vdots \\ L^j x_t &= x_{t-j} \end{aligned}$$

  Note that $L^{-j} x_t = x_{t+j}$.

- We can also define lag polynomials of the form

$$a(L) = \left(a_0 L^0 + a_1 L^1 + \cdots + a_p L^p\right) x_t = a_0 x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p}$$
$$a(L)b(L) = b(L)a(L),\ a(L)^2 = a(L)a(L)$$

- Example of ARMA(p,q) model:

$$\phi(L)x_t = \theta(L)\epsilon_t$$

- The product of two lag polynomials $a(L)$ and $b(L)$ is itself a lag polynomial $c(L) = a(L)b(L)$ of the form

$$\sum_{i=0}^{p} c_i L^i = \sum_{j=0}^{p} b_j L^j \sum_{k=1}^{p} a_k L^k$$

- In order to find these values analytically we can solve iteratively to obtain

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= a_0 b_1 + a_1 b_0 \\ c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \\ &\vdots \end{aligned}$$

- The inverse of a lag polynomial $a(L)$ is a polynomial $b(L) = a(L)^{-1}$ with the property $a(L)b(L) = 1$, and is of the form

$$b(L) = \left(a_0 L^0 + a_1 L^1 + \cdots + a_p L^p\right)^{-1}$$

- In order to find these values analytically we use the properties in the previous page and the fact that $c(L) = a(L)b(L) = 1$ to obtain

$$\begin{gathered} b_0 = a_0^{-1} \\ b_1 = -a_1/a_0^2 \\ b_2 = a_1^2/a_0^3 - a_2/a_0^2 \\ \vdots \end{gathered}$$