

Econometrics



David Preinerstorfer

Institute for Statistics and Mathematics
Dept. of Finance, Accounting and Statistics

Spring 2026



Last week:

1. Consistency.
2. Omitted variables, variable selection.
3. Regression diagnostics.
4. Taking care of heteroskedasticity and autocorrelation.

Plan for today:

1. Time series
2. Empirical autocovariances, stationarity testing, residual analysis
3. Models: Stationary processes, AR, MA, and ARMA models.

- Time series
- ARMA models
- GARCH models
- Non-stationary models

- Statistical Properties of OLS Estimation
- Specifications
- Regression diagnostics
- Time Series
- ARMA models

- A *times series* is a collection of observations $(y_1, \dots, y_T)$ made sequentially over time $t = 1, \dots, T$
- Observations are usually taken at equally spaced time points (frequency: yearly, monthly, quarterly, daily)
- Univariate time series: **single observation** at each time point t
- Multivariate time series: **multiple observations** at each time point t

SPY plotted against time:



QQQ plotted against time:



Gold price plotted against time:



BTC-USD plotted against time:



- Classical statistical methods assume that the data $y_1, \dots, y_T$ are realizations of a **sequence of iid random variables** Y_t .
- Time series analysis takes dependence between adjacent observations into account.
- Time series analysis assumes that the data $y_1, \dots, y_T$ are a realization of a **discrete-time stochastic process** $Y_t, t = 1, \dots, T$.
- Each observation y_t is a realized value of the random variable $Y_t, t = 1, \dots, T$. Time series observations y_t and y_s are **dependent**, if the corresponding random variable Y_t and Y_s are dependent.

The empirical autocorrelation function

The **empirical autocorrelation** r_1 at **lag 1** is defined as:

$$r_1 = \text{Corr}(y_{t-1}, y_t) = \frac{1}{T} \sum_{t=2}^T \left(\frac{y_{t-1} - \bar{y}}{s_y} \right) \cdot \left(\frac{y_t - \bar{y}}{s_y} \right) = \frac{\sum_{t=2}^T (y_{t-1} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

The correlation between time series observations h periods apart, i.e., y_t and y_{t-h} , is called the **empirical autocorrelation at lag h**

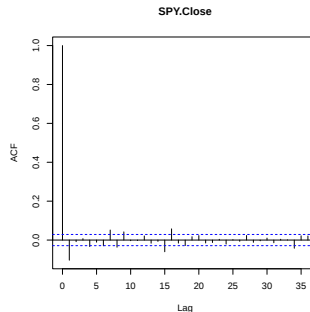
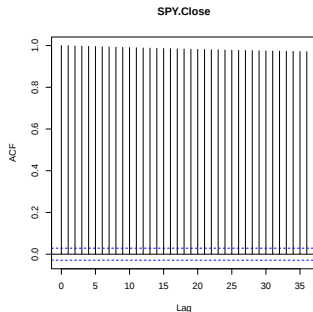
$$r_h = \text{Corr}(y_{t-h}, y_t) = \frac{1}{T} \sum_{t=h+1}^T \left(\frac{y_{t-h} - \bar{y}}{s_y} \right) \cdot \left(\frac{y_t - \bar{y}}{s_y} \right).$$

Note that $r_0 = 1$ and $-1 \leq r_h \leq 1$.

The empirical autocorrelogram

The plot of r_h against lag h is called the **correlogram**. It displays information about serial dependence in the time series (serial dependence over h periods small, if $r_h \approx 0$).

In R you use the function `acf` after converting a time series to a ts object (left unadjusted, right diff-log)



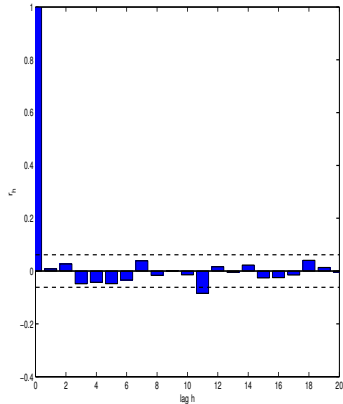
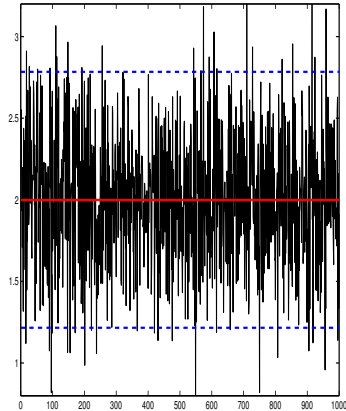
Testing in the empirical correlogram I

- The empirical autocorrelation $r_h = \text{Corr}(y_{t-h}, y_t)$ is an estimator for the theoretical autocorrelation $\rho(h) = \text{Corr}(Y_t, Y_{t-h})$ of the underlying process Y_t .
- If the underlying process Y_t is an iid sequence, then the autocorrelation $\rho(h)$ is 0 at any lag $h \neq 0$:

$$\rho(h) = \begin{cases} 1 & \text{for } h = 0 \\ 0 & \text{for } h = \pm 1, \pm 2, \dots \end{cases}$$

Even if the underlying process Y_t is iid, the empirical autocorrelations r_h for T realizations $y_1, \dots, y_T$ are typically different from 0 for $h \geq 1$ due to statistical variation:

Testing in the empirical correlogram II



Testing in the empirical correlogram III

If the underlying process Y_t is iid , then the empirical ACF r_h of any sequence of observations $y_t, t = 1, \dots, T$ has the following property:

$$E(r_h) = -\frac{1}{T}, \quad V(r_h) = \frac{1}{T}.$$

Asymptotically, i.e., for T large, r_h is normally distributed:

$$r_h \sim \text{Normal} \left(-\frac{1}{T}, \frac{1}{T} \right)$$

Test autocorrelations independently for each lag h .

- Under the null hypothesis $H_0 : \rho(h) = 0$, the absolute value of r_h does not exceed $2/\sqrt{T}$ with probability approximately equal to 0.95, i.e.,

$$P(|r_h| \leq \frac{2}{\sqrt{T}}) \approx 0.95.$$

Note: due to multiple testing, the risk to reject a true null hypothesis H_0 is larger than 5%.

Box-Ljung-Statistics

- A test statistic for the null hypothesis that there is no autocorrelation up to order h
 $H_0 : \rho(1) = \dots = \rho(h) = 0$ is the **Box-Ljung-Statistic** (Q -statistic) at lag h
- The Box-Ljung-Statistic does not assess single autocorrelations but aggregates all autocorrelations $r_1, \dots, r_h$ up to lag h

$$Q(h) = T(T+2) \sum_{k=1}^h \frac{r_k^2}{T-k}$$

where T is the number of observations.

- Under the null hypothesis (no autocorrelation up to lag h) Q is asymptotically distributed as χ^2 with h degrees of freedom

$$Q(h) \sim \chi_h^2$$

- The null hypothesis is **rejected**, if $Q(h) > \chi_{h;1-\alpha}^2$, or of the corresponding p -value is very small.

Both testing methods have little power for small T and might detect small significant, but irrelevant autocorrelations for T large.

Residual analysis in a regression model involving time series is important, because errors in such a model are often autocorrelated, i.e.,

$$\text{Cov}(\boldsymbol{\varepsilon}|\mathbf{X}) = \text{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{pmatrix}$$

Residual diagnostics

- Use the empirical ACF of the OLS residuals to test, if assumption **AH** (uncorrelated residuals) holds. Since the mean of the OLS residuals is equal to 0, the computation of r_h simplifies to:

$$r_h = \frac{1}{\epsilon' \epsilon} \sum_{i=h+1}^N \hat{\epsilon}_i \hat{\epsilon}_{i-h}.$$

- Check the Durbin-Watson Statistic (standard output of a regression analysis in most software packages).
- Perform the Breusch-Godfrey test

The Durbin-Watson Statistic

$$d = \frac{\sum_{i=2}^N (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^N \hat{\varepsilon}_i^2}$$

can be used to check for first-order serial correlation of the OLS residuals $\hat{\varepsilon}_i = y_i - \hat{y}_i$ in a regression model:

- d is an estimator of

$$D = \frac{E((\varepsilon_i - \varepsilon_{i-1})^2)}{E(\varepsilon_i^2)}.$$

If **AH** (*homoskedasticity*) holds (i.e., $V(\varepsilon_i^2) = \sigma^2$), then

$$D = \frac{E(\varepsilon_i^2) + E(\varepsilon_{i-1}^2) - 2E(\varepsilon_{i-1}\varepsilon_i)}{E(\varepsilon_i^2)} = \frac{\sigma^2 + \sigma^2}{\sigma^2} - 2\rho(1) = 2(1 - \rho(1)).$$

If there is no serial correlation in the errors ε_i at lag 1, i.e., $E(\varepsilon_{i-1}\varepsilon_i) = 0$, then $D = 2$. Hence, the Durbin-Watson statistic d will be close to 2. If there is serial correlation in the errors ε_i at lag 1, then $D = 2(1 - \rho(1))$ is different from 2.

- If autocorrelation at lag 1 is **positive**, i.e., $\rho(1) > 0$, then $0 \leq D < 2$. Hence, the Durbin-Watson statistic will be (considerably) smaller than 2.
- If autocorrelation at lag 1 is **negative**, i.e., $\rho(1) < 0$, then $2 \leq D \leq 4$. Hence, the Durbin-Watson statistic will be (considerably) larger than 2.
- No explicit sampling distribution is available for the Durbin-Watson statistic d .

Use the empirical ACF to test more formally the null hypothesis that no serial correlation is present.

The Breusch-Godfrey test

The *Breusch-Pagan test for autocorrelation* is based on an auxiliary regression of $\hat{\varepsilon}_i$ on p lagged residuals $\hat{\varepsilon}_{i-1}, \dots, \hat{\varepsilon}_{i-p}$, and the regressors:

$$\hat{\varepsilon}_i = \mathbf{x}_i\boldsymbol{\beta} + \gamma_1\hat{\varepsilon}_{i-1} + \dots + \gamma_p\hat{\varepsilon}_{i-p} + \xi_i. \quad (62)$$

Test the null hypothesis $H_0: \gamma_1 = \dots = \gamma_p = 0$ using the F-statistic (follows $F_{p, N-(K+p+1)}$ under H_0) or the test statistic $N \cdot R^2$, where R^2 is the coefficient of determination of the auxiliary regression model (62), which follows a χ_p^2 distribution under H_0 .

OLS estimation under autocorrelated errors I

- If the errors are correlated, then the standard errors of the OLS estimator are biased. E.g. if the majority of autocorrelations is positive, then standard errors are too small.
- The **Newey-West (HAC) estimator** corrects the standard errors of the OLS estimator. It requires the choice of a lag length p , such that all correlation ρ_j are set to 0, if $j > p$ and reduces to the White Heteroscedasticity consistent estimator for $p = 0$.
- More efficient estimator than OLS are available, if we know Ω (GLS estimator) or if we model the dependence structure in Ω , e.g., by an ARMA model (maximum likelihood estimation).

OLS estimation under autocorrelated errors II

Newey-West (HAC) estimator

It can also be applied, if the variance is heteroskedastic, i.e., $V(\varepsilon_i) = \sigma_i^2$, hence $\text{Cov}(\varepsilon_i, \varepsilon_{i-j}) = \sigma_i \sigma_{i-j} \rho_j$, or equivalently,

$$\text{Cov}(\varepsilon | \mathbf{X}) = \mathbf{\Sigma}^{1/2} \mathbf{\Omega} \mathbf{\Sigma}^{1/2}, \quad \mathbf{\Sigma} = \text{Diag}(\sigma_1^2 \dots \sigma_N^2).$$

In this case, the covariance matrix of the OLS estimator is given by:

$$\text{Cov}(\hat{\beta} | \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{\Omega} (\mathbf{X}'\mathbf{X})^{-1}, \quad (63)$$

where $\mathbf{\Omega} = \mathbf{X}' \mathbf{\Sigma}^{1/2} \mathbf{\Omega} \mathbf{\Sigma}^{1/2} \mathbf{X}$. (63) is robust to heteroskedastic and autocorrelated errors, hence the name **HAC**.

OLS estimation under autocorrelated errors III

Since $\sigma_{i,i-j} := \sigma_i \sigma_{i-j} \rho_j = 0$ for $j > p$, Ω is given by:

$$\Omega = \sum_{i=1}^N \sigma_i^2 \mathbf{x}_i' \mathbf{x}_i + \sum_{j=1}^p \sum_{i=j+1}^N \sigma_{i,i-j} (\mathbf{x}_i' \mathbf{x}_{i-j} + \mathbf{x}_{i-j}' \mathbf{x}_i) \approx$$
$$\frac{N}{N - (K + 1)} \left(\sum_{i=1}^N \hat{\varepsilon}_i^2 \mathbf{x}_i' \mathbf{x}_i + \sum_{j=1}^p w_j \sum_{i=j+1}^N \hat{\varepsilon}_i \hat{\varepsilon}_{i-j} (\mathbf{x}_i' \mathbf{x}_{i-j} + \mathbf{x}_{i-j}' \mathbf{x}_i) \right) =: \hat{\Omega}.$$

Note that $\hat{\varepsilon}_i \hat{\varepsilon}_{i-j}$ is a (one-point) estimator of $\sigma_{i,i-j}$ for $j < p$ with weight $w_j = 1 - j/(p + 1)$.

Using $\hat{\Omega}$ as an estimator of Ω , one obtains asymptotically robust t-tests, F-tests, confidence intervals, etc. (under suitable assumptions).

Generalized least square (ctd.)

The weighted least square estimator (61) introduced to deal with heteroskedasticity is a special case of the generalized least square (GLS) estimator which is more efficient than OLS, if the covariance matrix $\text{Cov}(\varepsilon | \mathbf{X}) = \mathbf{\Omega}$ is **known** (note that in this definition, $\mathbf{\Omega}$ is a full covariance matrix, rather than a correlation matrix).

Consider the Cholesky decomposition of the inverse matrix $\mathbf{\Omega}^{-1} = \mathbf{C}'\mathbf{C}$ (or equivalently, $\mathbf{\Omega} = \mathbf{C}^{-1}(\mathbf{C}')^{-1}$).

The matrix $\mathbf{C}$ is used to transform the original regression model such that the new disturbances satisfy **assumption AH**.

This yields:

$$\mathbf{C}\mathbf{y} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon} \Rightarrow \mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \quad (64)$$

$$\mathbf{E}(\boldsymbol{\varepsilon}^* \mid \mathbf{X}^*) = \mathbf{C} \cdot \mathbf{E}(\boldsymbol{\varepsilon} \mid \mathbf{X}^*) = \mathbf{0},$$

$$\text{Cov}(\boldsymbol{\varepsilon}^* \mid \mathbf{X}^*) = \mathbf{C} \cdot \text{Cov}(\boldsymbol{\varepsilon} \mid \mathbf{X}^*)\mathbf{C}' = \mathbf{C}\boldsymbol{\Omega}\mathbf{C}' = \mathbf{C}\mathbf{C}^{-1}(\mathbf{C}')^{-1}\mathbf{C}' = \mathbf{I}.$$

OLS estimation for the modified regression model (64) is the BLUE for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}, \quad (65)$$

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}. \quad (66)$$

Proof of (65):

$$\begin{aligned}\hat{\beta}_{\text{GLS}} &= ((\mathbf{X}^*)' \mathbf{X}^*)^{-1} (\mathbf{X}^*)' \mathbf{y}^* = (\mathbf{X}' \mathbf{C}' \mathbf{C} \mathbf{X})^{-1} \mathbf{X}' \mathbf{C}' \mathbf{C} \mathbf{y} \\ &= (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{y}.\end{aligned}$$

Proof of (66):

$$\begin{aligned}\text{Cov}(\hat{\beta}_{\text{GLS}}) &= (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \text{Cov}(\boldsymbol{\varepsilon} \mid \mathbf{X}) \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}.\end{aligned}$$

Financial time series: stock prices and indices, exchange rates, commodity prices (usually denoted by p_t). Often, the quantity of interest are the **log returns** y_t (continuous compounding) or simple returns r_t (discrete compounding) :

$$y_t = \log p_t - \log p_{t-1} = \log(p_t/p_{t-1}),$$

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1.$$

Log and simple returns are related as follows:

$$y_t = \log(1 + r_t), \quad r_t = \exp(y_t) - 1.$$

and differ with respect to second and higher order moments and the mean $\bar{r}$ is not equal to $\bar{y}$.

Stylized facts for financial returns

- Often, log returns are assumed to follow a normal distribution, but typical deviations are skewness and fat tails. Fat tails may be captured by the Student t distribution.
- Test for normality could be based on the histogram and the Jarque-Bera statistics or Quantile–Quantile plots.
- Financial returns are potentially autocorrelated, i.e., $\text{Corr}(y_t, y_{t-h}) \neq 0$ (consider empirical ACF of y_t) and display volatility clusters, i.e., $\text{Corr}(y_t^2, y_{t-h}^2) \neq 0$ or $\text{Corr}(|y_t|, |y_{t-h}|) \neq 0$ (consider empirical ACF of y_t^2 or $|y_t|$). This means that the returns y_t and y_{t-h} are dependent, even if they are uncorrelated.

Discrete-time Stochastic processes

- A times series $y_1, \dots, y_T$ is considered a realization of a discrete-time stochastic process Y_t , $t = 1, \dots, T$.
- A discrete-time stochastic processes can be described by specifying the joint probability distribution of $Y_{t_1}, \dots, Y_{t_n}$ for any set of time points $t_1, \dots, t_n$ and any value n
- For each $t = 1, \dots, T$, the process Y_t has
 - finite mean $\mu_t = E(Y_t)$ which defines the expected level and finite variance $\sigma_t^2 = V(Y_t)$ which describes the fluctuations of Y_t around μ_t .
- At time t , the process Y_t varies with probability $1 - \alpha$ between $[\mu_t - c_{t;\alpha/2}\sigma_t, \mu_t + c_{t;1-\alpha/2}\sigma_t]$, where $c_{t;p}$ is the appropriate quantile (e.g., of the standardized normal distribution, if Y_t is Gaussian).
- For mathematical simplicity, one often assumes underlying process Y_t with $t \in \mathbb{N}$ or $t \in \mathbb{Z}$.

The autocovariance function

$$\gamma(t, s) = E\left((Y_t - \mu_t)(Y_s - \mu_s)\right) = \gamma(s, t)$$

and the autocorrelation function (ACF)

$$\rho(t, s) = E\left(\frac{Y_t - \mu_t}{\sigma_t} \cdot \frac{Y_s - \mu_s}{\sigma_s}\right) = \frac{\gamma(t, s)}{\sigma_t \sigma_s} = \rho(s, t)$$

describe dependence between Y_t and Y_s . The random variables Y_t and Y_s at different time points t and s are possibly correlated, if $\gamma(t, s)$ and $\rho(t, s)$ are different from 0.

Autocovariance and autocorrelation function II

- If $\rho(t, s) = 0$, then Y_t and Y_s are uncorrelated.
- If $\rho(t, s) > 0$ ($t < s$), then Y_t and Y_s are positively correlated
 - for $Y_t > \mu_t$ we expect $Y_s > \mu_s$
 - for $Y_t < \mu_t$ we expect $Y_s < \mu_s$
- If $\rho(t, s) < 0$ ($t < s$), then Y_t and Y_s are negatively correlated
 - for $Y_t > \mu_t$ we expect $Y_s < \mu_s$
 - for $Y_t < \mu_t$ we expect $Y_s > \mu_s$

Dependence between Y_t and $Y_{t-1}, Y_{t-2}, \dots$ implies that the conditional distribution of $Y_t \mid I_{t-1}$, where $I_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$ refers to the information available at time $t - 1$, is different from the marginal distribution of Y_t :

$$E(Y_t \mid I_{t-1}) \neq \mu_t = E(Y_t)$$

$$V(Y_t \mid I_{t-1}) \neq \sigma_t^2 = V(Y_t).$$

Time series models are usually defined as dynamic stochastic models, by specifying the conditional distribution of $Y_t \mid I_{t-1}$.

Dynamic stochastic models

Specify the conditional expectation $E(Y_t | I_{t-1}, \theta)$ and the conditional variance $V(Y_t | I_{t-1}, \theta)$ up to unknown parameters θ .

If the conditional distribution of $Y_t | I_{t-1}, \theta$ is assumed to be a well-known distribution family (e.g., normal distribution) with p.d.f. $p(y_t | I_{t-1}, \theta)$, then this defines the p.d.f. $p(y_1, \dots, y_T | \theta, I_0)$ of the joint distribution of $Y_1, \dots, Y_T$ given θ and initial information I_0 :

$$p(y_1, \dots, y_T | \theta, I_0) = \prod_{t=1}^T p(y_t | I_{t-1}, \theta). \quad (67)$$

Properties of the joint as well as the marginal distribution $p(y_t | \theta, I_0)$ can be derived from the conditional distribution.

White noise processes

A **white noise process** Y_t has a constant mean (i.e., $E(Y_t) \equiv \mu$) and is a sequence of uncorrelated random variables, i.e.,

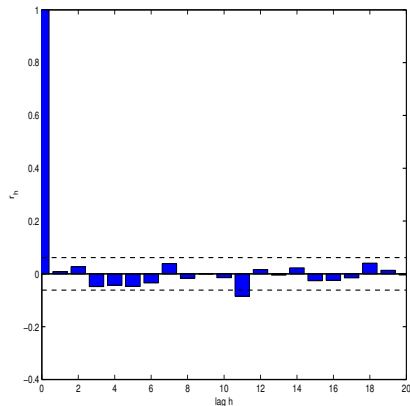
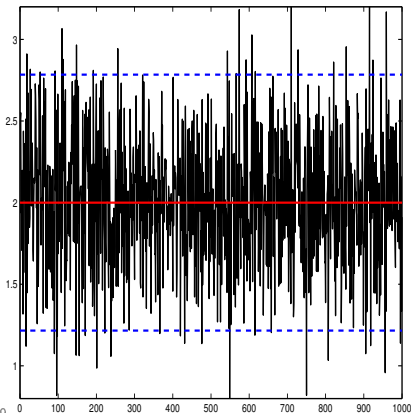
$$\gamma(h) = \rho(h) = 0 \quad \text{for } h = \pm 1, \pm 2, \dots$$

For a **zero mean white noise process**, $\mu = 0$.

A Gaussianity assumption here implies for a white noise process that the random variables Y_t are an **independent white noise process** (but not necessarily identically distributed, in particular $V(Y_t)$ need not be constant).

Example 1: Gaussian iid sequence

$Y_t = 2 + \varepsilon_t$, $\varepsilon_t \sim \text{Normal}(0, 0.16)$ Left hand side: simulated path of Y_t (Red/full line: $\mu_t \equiv 2$, $\sigma_t^2 \equiv 0.16$; blue/dashed lines: $\mu \pm 1.96\sigma = 2 \pm 1.96 \cdot 0.4$); right hand side: empirical ACF



A stochastic processes is called **weakly stationary** (second order stationary), if the second moments of Y_t exist and

- mean and variance are constant over time: $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$:
- the autocovariance function $\gamma(t, s)$ and the autocorrelation function $\rho(t, s)$ depend only on the *lag* $h = t - s$, i.e.,

$$\gamma(t, s) = \gamma(t - s) = \gamma(h), \quad \rho(t, s) = \rho(t - s) = \rho(h).$$

Strong stationarity: for any value $1 \leq n \leq T$, any $t_1 < t_2 < \dots < t_n$, and any integer h the joint probability distributions of $Y_{t_1}, \dots, Y_{t_n}$ and $Y_{t_1+h}, \dots, Y_{t_n+h}$ are the same (assuming that the indexation makes sense).

Mean reverting processes

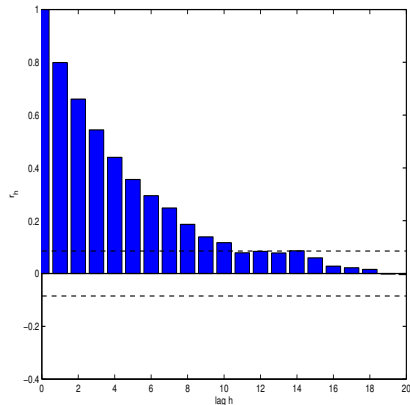
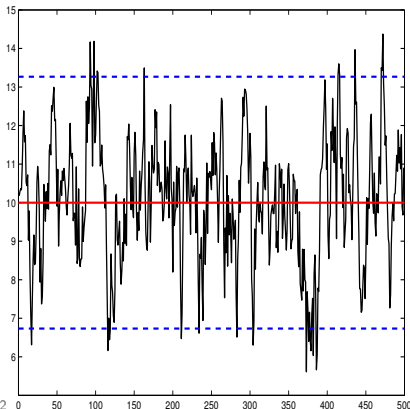
Stationary processes are typically mean reverting:

- If $\rho(h) > 0$, then Y_t tends to be above (below) the long-run level μ , if Y_{t-h} is above (below) the long-run level μ .
- If $\rho(h) < 0$, then Y_t tends to be above (below) the long-run level μ , if Y_{t-h} is below (above) the long-run level μ .

In both cases, the process reverts to the level μ and is independent of the current value Y_t in the long run.

Example 2: AR(1)-process

$Y_t = 0.8Y_{t-1} + 2 + \varepsilon_t, \quad \varepsilon_t \sim \text{Normal}(0, 1)$ Left hand side: simulated path of Y_t (Red/full line: $\mu_t \equiv 10, \sigma_t^2 \equiv 1/0.36$; blue/dashed lines: $\mu \pm 1.96\sigma = 10 \pm 1.96 \cdot 1/0.6$, right hand side: empirical ACF



Examples of non-stationary stochastic processes I

- The process $Y_t = Y_{t-1} + \varepsilon_t$ is called a **random walk** (ε_t is a white noise process). The random walk is not stationary, see Example 3.
- The process $Y_t = 2 + 0.01t + \varepsilon_t$, has an increasing level $\mu_t = 2 + 0.01t$, hence Y_t is not stationary, see Example 4. However, the process $Y_t - \mu_t = \varepsilon_t$ is stationary. Y_t is a **trend stationary random process**.

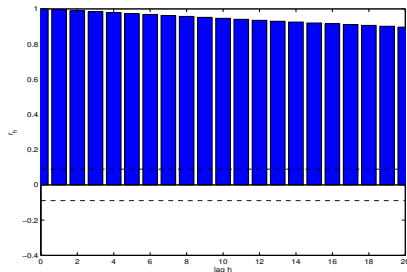
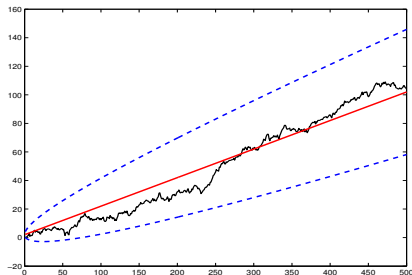
The empirical autocorrelogram (based on constant mean and variance) declines slowly, because both processes are non-stationary.

Examples of non-stationary stochastic processes II

Example 3: Random walk process with drift

$$Y_t = Y_{t-1} + 0.2 + \varepsilon_t, \quad \varepsilon_t \sim \text{Normal}(0, 1), \quad Y_0 = 0$$

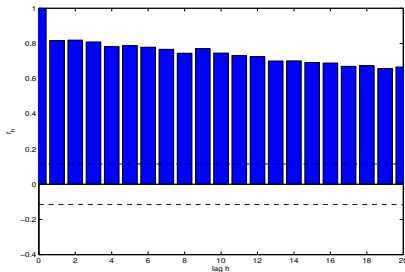
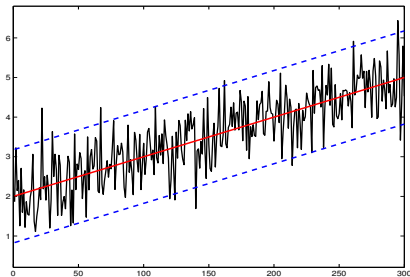
Left hand side: simulated path of Y_t (Red/full line: $\mu_t \equiv 0.2t$, $\sigma_t^2 = t$; blue/dashed lines: $\pm 1.96\sqrt{t}$), right hand side: empirical ACF



Example 4: trend stationary random process

$$Y_t = 2 + 0.01t + \varepsilon_t, \quad \varepsilon_t \sim \text{Normal}(0, 0.36)$$

Left hand side: simulated path of Y_t (Red/full line: $\mu_t = 2 + t$, $\sigma_t^2 \equiv 0.36$; blue/dashed lines: $\pm 1.96 \cdot 0.6$), right hand side: empirical ACF



Modeling stationary time series:

- ARMA models and GARCH models

Modeling non-stationary time series:

- Random walk processes and ARIMA models
- Trend stationarity: $\varepsilon_t = Y_t - \mu_t$ (where μ_t models a trend) is stationary

Unit root tests to test for stationarity

- Statistical Properties of OLS Estimation
- Specifications
- Regression diagnostics
- Time Series
- ARMA models

ARMA models

- Autoregressive moving-average (ARMA) processes model the conditional expectation $E(Y_t | I_{t-1}, \theta)$ as a dynamic model, whereas the conditional variance $V(Y_t | I_{t-1}, \theta) = \sigma_\varepsilon^2$ is assumed to be fixed (but unknown).
- Appropriate for modelling the autocorrelation structure of **stationary** time series.
- Stationary ARMA processes are completely identified from their autocovariance or autocorrelation structure $\gamma(h)$ and $\rho(h)$.

ARMA processes are dynamic stochastic models, driven by innovations (shocks) following a zero mean white noise process ε_t with constant variance σ_ε^2 , which is assumed to be independent of I_{t-1} , i.e.:

$$E(\varepsilon_t | I_{t-1}) = E(\varepsilon_t) = 0,$$

$$V(\varepsilon_t | I_{t-1}) = V(\varepsilon_t) = \sigma_\varepsilon^2.$$

There are basically two lines of thinking to introduce dependence of $E(Y_t | I_{t-1}, \theta)$ on the past information: **AR models** include past levels, e.g., Y_{t-j} , and **MA models** introduce past innovations ε_{t-j} . ARMA models combine both.

Autoregressive Processes I

The **AR(1)**-process is defined as

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad (68)$$

where ϕ is an unknown parameter and $t \in \mathbb{Z}$, if the process starts in the infinite past, and $t \in \mathbb{N}_0$, if the process starts with initial value Y_0 .

An AR(1) process is **stationary**, if $|\phi| < 1$.

It reduces to a zero-mean white noise process, if $\phi = 0$.

It corresponds to the random walk $Y_t = Y_{t-1} + \varepsilon_t$ for $\phi = 1$.

For $|\phi| > 1$, the process is “explosive”.

Moments of a stationary AR(1) process

The AR(1) process can be represented in the following form:

$$Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \quad (69)$$

The proof of (69) is based on recursive substitution:

$$\begin{aligned} Y_t &= \varepsilon_t + \phi Y_{t-1} = \varepsilon_t + \phi(\varepsilon_{t-1} + \phi Y_{t-2}) \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2(\varepsilon_{t-2} + \phi Y_{t-3}) = \cdots = \\ &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \end{aligned}$$

Autoregressive Processes III

If $|\phi| < 1$, then the absolutely summability condition holds:

$$\sum_{j=0}^{\infty} |\phi|^j = \frac{1}{1 - |\phi|},$$

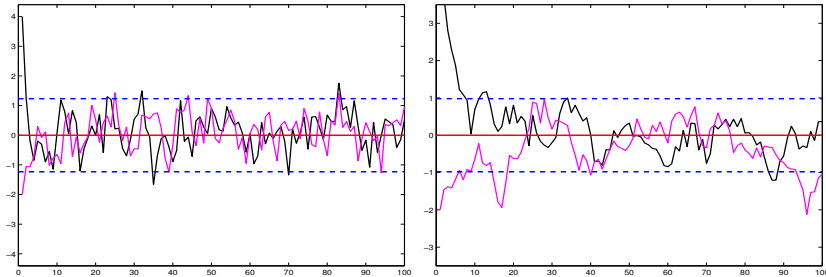
and the AR(1)-process is stationary:

$$E(Y_t) = E\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \phi^j E(\varepsilon_{t-j}) = 0,$$

$$V(Y_t) = V\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \phi^{2j} V(\varepsilon_{t-j}) = \frac{\sigma_{\varepsilon}^2}{1 - \phi^2}.$$

Examples of stationary AR(1)-Processes

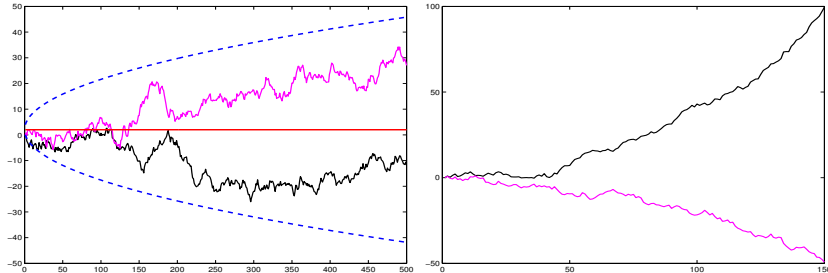
A stationary AR(1)-process is mean reverting, even if it starts from an extreme deviation from the mean: 2 simulated time series with $Y_0 = 4$ and $Y_0 = -2$; left: $\phi = 0.3$, $\sigma_\varepsilon^2 = 0.36$; right: $\phi = 0.8$; $\sigma_\varepsilon^2 = 0.09$)



Examples of nonstationary AR(1)-Processes

A non-stationary AR(1)-process is not mean reverting

left: $\phi = 1$ (random walk), right: $\phi = 1.02$ (explosive)



Autocorrelation of a stationary AR(1) process

For a stationary AR(1) process the autocorrelation function at lag $h > 0$ is given by $\rho(h) = \phi^h$.

Proof: use (69) and the uncorrelatedness of ε_t to show that

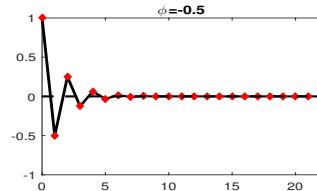
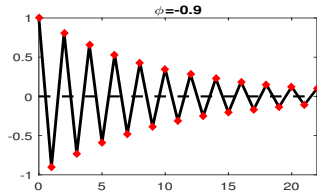
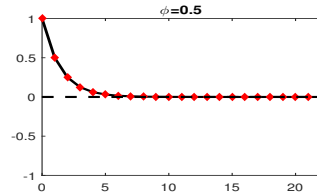
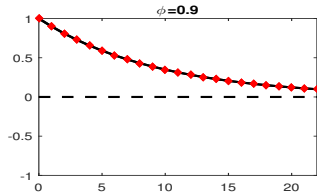
$$\begin{aligned}\text{Cov}(Y_t, Y_{t+h}) &= \text{E}(Y_t \cdot Y_{t+h}) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi^j \phi^k \text{E}(\varepsilon_{t-j} \varepsilon_{t+h-k}) = \\ &= \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \text{E}(\varepsilon_{t-j}^2) = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \phi^j \phi^{h+j} = \phi^h \frac{\sigma_{\varepsilon}^2}{1 - \phi^2} = \phi^h \sigma^2.\end{aligned}$$

Hence, $\text{Cov}(Y_t, Y_{t+h}) \equiv \gamma(h)$ for all t and $\rho(h) = \gamma(h)/\sigma^2 = \phi^h$.

The ACF of an AR(1) process **decays exponentially** toward 0:

- If $|\phi|$ is close to 0, then the ACF decays **quickly**.
- If $|\phi|$ is close to 1, then the ACF decays **slowly**.
- If $\phi < 0$, then the ACF **alternates** between positive and negative values and the process is oscillating between positive and negative values.

Autoregressive Processes VIII



ACF of an AR(1) process with various values of ϕ .

Y_t is called an **autoregressive process of order p or AR(p)-process** if

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t. \quad (70)$$

- p is the **order of the process**; $\phi_1, \dots, \phi_p$ are the parameters of the model

An AR(p) process is not necessarily stationary. The stationarity conditions are more complicated than for an AR(1)-process (more about this later).

Moving Average Processes I

The second way to construct correlated processes is to include past shocks ε_{t-j} into the definition of the conditional expectation.

The **MA(1) Process** Y_t is defined as

$$Y_t = \theta\varepsilon_{t-1} + \varepsilon_t.$$

For $\theta = 0$, the processes reduces to a zero-mean white noise process. The **moving average process of order q** or **MA(q)** process Y_t is defined as

$$Y_t = \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t. \quad (71)$$

q is the order of the model and $\theta_1, \dots, \theta_q$ are unknown parameters.

Moving Average Processes II

Moments of the MA processes

MA processes are always stationary (define $\theta_0 = 1$):

$$E(Y_t) = E\left(\sum_{j=0}^q \theta_j \varepsilon_{t-j}\right) = \sum_{j=0}^q \theta_j E(\varepsilon_{t-j}) = 0,$$

$$V(Y_t) = V\left(\sum_{j=0}^q \theta_j \varepsilon_{t-j}\right) = \sum_{j=0}^q \theta_j^2 V(\varepsilon_{t-j}) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2,$$

$$\rho(h) = \left(\sum_{j=0}^{q-h} \theta_j \theta_{j+h}\right) / \left(\sum_{j=0}^q \theta_j^2\right), \quad \text{for } 1 \leq h \leq q, \quad (72)$$

and $\rho(h) = \gamma(h) = 0$ for $h > q$. The ACF of the MA(q) process 'cuts off' at lag q .

Autocorrelation function

Proof of (72): use the uncorrelatedness of ε_t and define $\theta_j = 0$, if $j > q$:

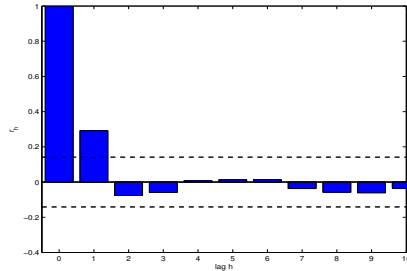
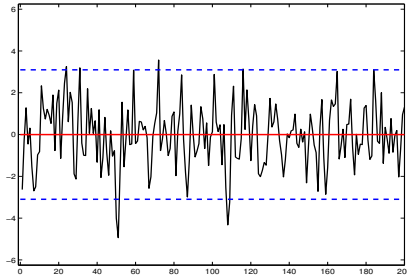
$$\begin{aligned}\text{Cov}(Y_t, Y_{t+h}) &= \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \mathbb{E}(\varepsilon_{t-j} \varepsilon_{t+h-k}) = \\ &= \sum_{j=0}^q \theta_j \theta_{j+h} \mathbb{E}(\varepsilon_{t-j}^2) = \sigma_\varepsilon^2 \left(\sum_{j=0}^{q-h} \theta_j \theta_{j+h} \right)\end{aligned}$$

does not depend on t : $\text{Cov}(Y_t, Y_{t+h}) = \gamma(h)$. Furthermore, $\gamma(h) = \rho(h) = 0$ for $h > q$.
Finally, $\rho(h) = \gamma(h)/\sigma^2$ for $h \leq q$.

For an **MA(1)** process is $\rho(1) = \theta/(1 + \theta^2)$ and 0 otherwise.

Example: MA(1)-Process with $\theta = 0.5$

Left: 200 time series observations, right: empirical autocorrelation function



Invertibility

The MA(1) process is not uniquely determined by its ACF as the processes

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad X_t = \varepsilon_t + \frac{1}{\theta} \varepsilon_{t-1}$$

have identical ACF:

$$\rho_X(1) = \frac{\frac{1}{\theta}}{1 + \frac{1}{\theta^2}} = \frac{\theta}{1 + \theta^2} = \rho_Y(1).$$

The MA(1) process is uniquely determined by its ACF if θ satisfies the **invertibility condition** $|\theta| < 1$.

Similar issues arise for general MA processes.