

In the previous chapters we have seen in which way the mean μ , and, more importantly, the autocovariance function, $\gamma(h)$, $h = 0, \pm 1, \pm 2, \dots$, of a stationary stochastic process $\{X_t\}$ characterize its dynamic properties, at least if we restrict ourselves to the first two moments. In particular, we have investigated how the autocovariance function is related to the coefficients of the corresponding ARMA process. Thus the estimation of the ACF is not only interesting for its own sake, but also for the specification and identification of appropriate ARMA models. It is therefore of outmost importance to have reliable (consistent) estimators for these entities. Moreover, we want to test specific features for a given time series. This means that we have to develop corresponding testing theory. As the small sample distributions are hard to get, we rely for this purpose on asymptotic theory.¹

In this section we will assume that the process is stationary and observed for the time periods $t = 1, 2, \dots, T$. We will refer to T as the sample size. As mentioned previously, the standard sampling theory is not appropriate in the time series context because the X_t 's are not independent draws from some underlying distribution, but are systematically related to each other.

4.1 Estimation of the Mean

The arithmetic average constitutes a “natural” estimator of the mean μ of the stochastic process. The arithmetic mean \bar{X}_T is defined as usual by

$$\bar{X}_T = \frac{1}{T} (X_1 + X_2 + \dots + X_T).$$

¹Recently, bootstrap methods have also been introduced in the time series context.

It is immediately clear that the arithmetic average is an unbiased estimator of the mean:

$$\mathbb{E}\bar{X}_T = \frac{1}{T} (\mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_T) = \mu.$$

Of greater interest are the asymptotic properties of the variance of the arithmetic mean $\mathbb{V}\bar{X}_T$ which are summarized in the following theorem:

Theorem 4.1 (Convergence of Arithmetic Average). *If $\{X_t\}$ is a stationary stochastic process with mean μ and ACF $\gamma(h)$ then the variance of the arithmetic mean $\mathbb{V}\bar{X}_T$ has the following asymptotic properties:*

$$\begin{aligned} \mathbb{V}\bar{X}_T &= \mathbb{E}(\bar{X}_T - \mu)^2 \rightarrow 0, & \text{if } \gamma(T) \rightarrow 0; \\ T\mathbb{V}\bar{X}_T &= T\mathbb{E}(\bar{X}_T - \mu)^2 \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h), & \text{if } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \end{aligned}$$

for T going to infinity.

Proof. Immediate algebra establishes:

$$\begin{aligned} 0 \leq T\mathbb{V}\bar{X}_T &= \frac{1}{T} \sum_{i,j=1}^T \text{cov}(X_i, X_j) = \sum_{|h|<T} \left(1 - \frac{|h|}{T}\right) \gamma(h) \\ &\leq \sum_{|h|<T} |\gamma(h)| = 2 \sum_{h=1}^T |\gamma(h)| + \gamma(0). \end{aligned}$$

The assumption $\gamma(h) \rightarrow 0$ for $h \rightarrow \infty$ implies that for any given $\varepsilon > 0$, we can find T_0 such that $|\gamma(h)| < \varepsilon/2$ for $h \geq T_0$. If $T > T_0$ and $T > 2T_0\gamma(0)/\varepsilon$ then

$$\begin{aligned} 0 \leq \frac{1}{T} \sum_{h=1}^T |\gamma(h)| &= \frac{1}{T} \sum_{h=1}^{T_0-1} |\gamma(h)| + \frac{1}{T} \sum_{h=T_0}^T |\gamma(h)| \\ &\leq \frac{T_0\gamma(0)}{T} + \frac{1}{T}(T - T_0)\varepsilon/2 \leq \frac{T_0\gamma(0)}{T} + \varepsilon/2 \leq \frac{T_0\gamma(0)\varepsilon}{2T_0\gamma(0)} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore $\mathbb{V}\bar{X}_T$ converges to zero for $T \rightarrow \infty$ which establishes the first property. Moreover, we have

$$\lim_{T \rightarrow \infty} T\mathbb{V}\bar{X}_T = \lim_{T \rightarrow \infty} \sum_{|h|<T} \left(1 - \frac{|h|}{T}\right) \gamma(h) = \sum_{h=-\infty}^{\infty} \gamma(h) < \infty.$$

The infinite sum $\sum_{h=-\infty}^{\infty} \gamma(h)$ converges because it converges absolutely by assumption. \square

This Theorem establishes that the arithmetic average is not only an unbiased estimator of the mean, but also a consistent one. In particular, the arithmetic average converges in the mean-square sense, and therefore also in probability, to the true mean (see appendix C). This result can be interpreted as a reflection of the concept of ergodicity (see Sect. 1.2). The assumptions are relatively mild and are fulfilled for the ARMA processes because for these processes $\gamma(h)$ converges exponentially fast to zero (see Sect. 2.4.2, in particular Eq. (2.6)). Under little more restrictive assumptions it is even possible to show that the arithmetic mean is asymptotically normally distributed.

Theorem 4.2 (Asymptotic Distribution of Sample Mean). *For any stationary process $\{X_t\}$ given by*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad Z_t \sim \text{IID}(0, \sigma^2),$$

such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, the arithmetic average \bar{X}_T is asymptotically normal:

$$\begin{aligned} \sqrt{T}(\bar{X}_T - \mu) &\xrightarrow{d} N\left(0, \sum_{h=-\infty}^{\infty} \gamma(h)\right) \\ &= N\left(0, \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2\right) = N(0, \sigma^2 \Psi(1)^2) \end{aligned}$$

where γ is the autocovariance function of $\{X_t\}$.

Proof. The standard proof invokes the Basic Approximation Theorem C.14 and the Central Limit Theorem for m -dependent processes C.13. To this end we define the $2m$ -dependent approximate process

$$X_t^{(m)} = \mu + \sum_{j=-m}^m \psi_j Z_{t-j}.$$

For $\{X_t^{(m)}\}$, we have $V_m = \sum_{h=-m}^m \gamma(h) = \sigma^2 (\sum_{j=-m}^m \psi_j)^2$. This last assertion can be verified by noting that

$$\begin{aligned} V &= \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \\ &= \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \sum_{h=-\infty}^{\infty} \psi_{j+h} = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2. \end{aligned}$$

Note that the assumption $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ guarantees the convergence of the infinite sums. Applying this result to the special case $\psi_j = 0$ for $|j| > m$, we obtain V_m .

The arithmetic average of the approximating process is

$$\bar{X}_T^{(m)} = \frac{1}{T} \sum_{t=1}^T X_t^{(m)}.$$

The CLT for m -dependent processes C.13 then implies that for $T \rightarrow \infty$

$$\sqrt{T} \left(\bar{X}_T^{(m)} - \mu \right) \xrightarrow{d} X^{(m)} = N(0, V_m).$$

As $m \rightarrow \infty$, $\sigma^2(\sum_{j=-m}^m \psi_j)^2$ converges to $\sigma^2(\sum_{j=-\infty}^{\infty} \psi_j)^2$ and thus

$$X^{(m)} \xrightarrow{d} X = N(0, V) = N\left(0, \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2\right).$$

This assertion can be established by noting that the characteristic functions of $X^{(m)}$ approaches the characteristic function of X so that by Theorem C.11 $X^{(m)} \xrightarrow{d} X$.

Finally, we show that the approximation error becomes negligible as T goes to infinity:

$$\sqrt{T} (\bar{X}_T - \mu) - \sqrt{T} (\bar{X}_T^{(m)} - \mu) = T^{-1/2} \sum_{t=1}^T (X_t - X_t^{(m)}) = T^{-1/2} \sum_{t=1}^T e_t^{(m)}$$

where the error $e_t^{(m)}$ is

$$e_t^{(m)} = \sum_{|j|>m} \psi_j Z_{t-j}.$$

Clearly, $\{e_t^{(m)}\}$ is a stationary process with autocovariance function γ_e such that $\sum_{h=-\infty}^{\infty} \gamma_e(h) = \sigma^2 \left(\sum_{|j|>m} \psi_j \right)^2 < \infty$. We can therefore invoke Theorem 4.1 to show that

$$\mathbb{V} \left(\sqrt{T} (\bar{X}_T - \mu) - \sqrt{T} (\bar{X}_T^{(m)} - \mu) \right) = T \mathbb{V} \left(\frac{1}{T} \sum_{t=1}^T e_t^{(m)} \right)$$

converges to $\sigma^2 \left(\sum_{|j|>m} \psi_j \right)^2$ as $T \rightarrow \infty$. This term converges to zero as $m \rightarrow \infty$. The approximation error $\sqrt{T}(\bar{X}_T - \mu) - \sqrt{T}(\bar{X}_T^{(m)} - \mu)$ therefore converges in mean square to zero and thus, using Chebyshev's inequality (see Theorem C.3 or C.7), also in probability. We have therefore established the third condition of Theorem C.14 as well. Thus, we can conclude that $\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{d} X$. \square

Under a more restrictive summability condition which holds, however, within the context of causal ARMA processes, we can provide a less technical proof. This proof follows an idea of Phillips and Solo (1992) and is based on the Beveridge-Nelson decomposition (see Appendix D).²

Theorem 4.3. *For any stationary process*

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

with the properties $Z_t \sim \text{IID}(0, \sigma^2)$ and $\sum_{j=0}^{\infty} j^2 |\psi_j|^2 < \infty$, the arithmetic average \bar{X}_T is asymptotically normal:

$$\begin{aligned} \sqrt{T}(\bar{X}_T - \mu) &\xrightarrow{d} N\left(0, \sum_{h=-\infty}^{\infty} \gamma(h)\right) \\ &= N\left(0, \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j\right)^2\right) = N(0, \sigma^2 \Psi(1)^2). \end{aligned}$$

Proof. The application of the Beveridge-Nelson decomposition (see Theorem D.1 in Appendix D) leads to

$$\begin{aligned} \bar{X}_T - \mu &= \frac{1}{T} \sum_{t=1}^T \Psi(L)Z_t = \frac{1}{T} \sum_{t=1}^T (\Psi(1) - (L-1))\tilde{\Psi}(L)Z_t \\ &= \Psi(1) \left(\frac{1}{T} \sum_{t=1}^T Z_t \right) + \frac{1}{T} \tilde{\Psi}(L)(Z_0 - Z_T) \\ \sqrt{T}(\bar{X}_T - \mu) &= \Psi(1) \left(\sqrt{T} \frac{\sum_{t=1}^T Z_t}{T} \right) + \frac{1}{\sqrt{T}} \tilde{\Psi}(L)Z_0 - \frac{1}{\sqrt{T}} \tilde{\Psi}(L)Z_T. \end{aligned}$$

²The Beveridge-Nelson decomposition is an indispensable tool for the understanding of integrated and cointegrated processes analyzed in Chaps. 7 and 16.

The assumption $Z_t \sim \text{IID}(0, \sigma^2)$ allows to invoke the Central Limit Theorem C.12 of Appendix C to the first term. Thus, $\sqrt{T} \frac{\sum_{t=1}^T Z_t}{T}$ is asymptotical normal with mean zero and variance σ^2 . Theorem D.1 also implies $|\Psi(1)| < \infty$. Therefore, the term $\Psi(1)\sqrt{T} \frac{\sum_{t=1}^T Z_t}{T}$ is asymptotically normal with mean zero and variance $\sigma^2\Psi(1)^2$.

The variances of the second and third term are equal to $\frac{\sigma^2}{T} \sum_{j=0}^T \tilde{\psi}_j^2$. The summability condition then implies according to Theorem D.1 that $\sum_{j=0}^T \tilde{\psi}_j^2$ converges for $T \rightarrow \infty$. Thus, the variances of the last two terms converge to zero implying that these terms converge also to zero in probability (see Theorem C.7) and thus also in distribution. We can then invoke Theorem C.10 to establish the Theorem. Finally, the equality of $\sum_{h=-\infty}^{\infty} \gamma(h)$ and $\sigma^2\Psi(1)^2$ can be obtained from direct computations or by the application of Theorem 6.4. \square

Remark 4.1. Theorem 4.2 holds with respect to any causal ARMA process because the ψ_j 's converge exponentially fast to zero (see the discussion following Eq. (2.5)).

Remark 4.2. If $\{X_t\}$ is a Gaussian process, then for any given fixed T , \bar{X}_T is distributed as

$$\sqrt{T}(\bar{X}_T - \mu) \sim \text{N}\left(0, \sum_{|h| < T} \left(1 - \frac{|h|}{T}\right) \gamma(h)\right).$$

According to Theorem 4.2, the asymptotic variance of the average depends on the sum of all covariances $\gamma(h)$. This entity, denoted by J , is called the *long-run variance* of $\{X_t\}$:

$$J = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) \left(1 + 2 \sum_{h=1}^{\infty} \rho(h)\right). \quad (4.1)$$

Note that the long-run variance equals 2π times the spectral density $f(\lambda)$ evaluated at $\lambda = 0$ (see the Definition 6.1 of the spectral density in Sect. 6.1).

As the long-run variance takes into account the serial properties of the time series, it is also called heteroskedastic and autocorrelation consistent variance (HAC variance). If $\{X_t\}$ has some nontrivial autocorrelation (i.e. $\rho(h) \neq 0$ for $h \neq 0$), the long-run variance J is different from $\gamma(0)$. This implies among other things that the construction of the t-statistic for testing the simple hypothesis $H_0: \mu = \mu_0$ should be based on J rather than on $\gamma(0)$.

In case that $\{X_t\}$ is a causal ARMA process with $\Phi(L)X_t = \Theta(L)Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$, the long-run variance is given by

$$J = \left(\frac{\Theta(1)}{\Phi(1)}\right)^2 \sigma^2 = \Psi(1)^2 \sigma^2.$$

If $\{X_t\}$ is a AR(1) process with $X_t = \phi X_{t-1} + Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$ and $|\phi| < 1$, $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$ and $\rho(h) = \phi^{|h|}$. Thus the long-run variance is given by $J = \frac{\sigma^2}{(1-\phi)^2} = \gamma(0) \times \frac{1+\phi}{1-\phi}$. From this example it is clear that the long-run variance can be smaller or larger than $\gamma(0)$, depending on the sign of ϕ : for negative values of ϕ , $\gamma(0)$ overestimates the long-run variance; for positive values, it underestimates J . The estimation of the long-run variance is dealt with in Sect. 4.4.

4.2 Estimation of the Autocovariance and the Autocorrelation Function

With some slight, asymptotically unimportant modifications, we can use the standard estimators for the autocovariances, $\gamma(h)$, and the autocorrelations, $\rho(h)$, of a stationary stochastic process:

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T), \quad (4.2)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \quad (4.3)$$

These estimators are biased because the sums are normalized (divided) by T rather than $T - h$. The normalization with $T - h$ delivers an unbiased estimate only if \bar{X}_T is replaced by μ which, however is typically unknown in practice. The second modification concerns the use of the complete sample for the estimation of μ .³ The main advantage of using the above estimators is that the implied estimator for the covariance matrix, $\hat{\Gamma}_T$, respectively the autocorrelation matrix, \hat{R}_T , of $(X_1, \dots, X_T)'$,

$$\hat{\Gamma}_T = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(T-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(T-1) & \hat{\gamma}(T-2) & \dots & \hat{\gamma}(0) \end{pmatrix}$$

$$\hat{R}_T = \frac{\hat{\Gamma}_T}{\hat{\gamma}(0)}$$

always delivers, independently of the realized observations, non-negative definite and for $\hat{\gamma}(0) > 0$ non-singular matrices. The resulting estimated autocovariance function will then satisfy the characterization given in Theorem 1.1, in particular property (iv).

³The standard statistical formulas would suggest to estimate the mean appearing in first multiplicand from X_1, \dots, X_{T-h} , and the mean appearing in the second multiplicand from X_{h+1}, \dots, X_T .

According to Box and Jenkins (1976, p. 33), one can expect reasonable estimates for $\gamma(h)$ and $\rho(h)$ if the sample size is larger than 50 and if the order of the autocorrelation coefficient is smaller than $T/4$.

The theorem below establishes that these estimators lead under rather mild conditions to consistent and asymptotically normally distributed estimators.

Theorem 4.4 (Asymptotic Distribution of Autocorrelations). *Let $\{X_t\}$ be the stationary process*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

with $Z_t \sim \text{IID}(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} j |\psi_j|^2 < \infty$. Then we have for $h = 1, 2, \dots$

$$\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \xrightarrow{d} \text{N} \left(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix}, \frac{W}{T} \right)$$

where the elements of $W = (w_{ij})_{i,j \in \{1, \dots, h\}}$ are given by Bartlett's formula

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)][\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)].$$

Proof. Brockwell and Davis (1991, section 7.3) □

Brockwell and Davis (1991) offer a second version of the above theorem where $\sum_{j=-\infty}^{\infty} j |\psi_j|^2 < \infty$ is replaced by the assumption of finite fourth moments, i.e. by assuming $\mathbb{E}Z_t^4 < \infty$. As we rely mainly on ARMA processes, we do not pursue this distinction further because this class of process automatically fulfills the above assumptions as soon as $\{Z_t\}$ is identically and independently distributed (IID). A proof which relies on the Beveridge-Nelson polynomial decomposition (see Theorem D.1 in Appendix D) can be gathered from Phillips and Solo (1992).

Example: $\{X_t\} \sim \text{IID}(0, \sigma^2)$

The most important application of Theorem 4.4 is related to the case of a white noise process. For this process $\rho(h)$ is equal to zero for $|h| > 0$. Theorem 4.4 then implies that

$$w_{ij} = \begin{cases} 1, & \text{for } i = j; \\ 0, & \text{otherwise.} \end{cases}$$

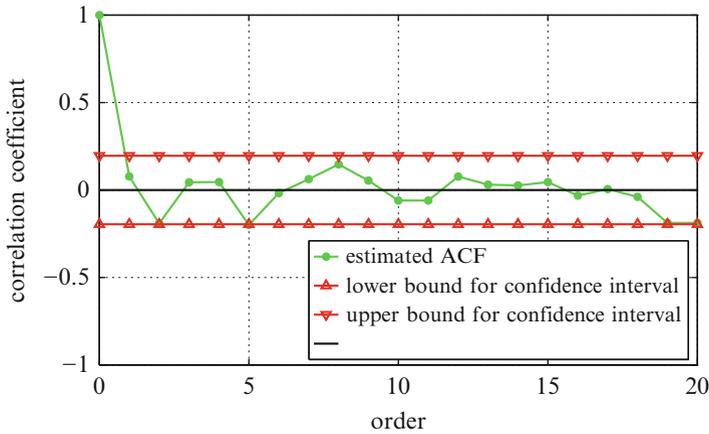


Fig. 4.1 Estimated autocorrelation function of a WN(0,1) process with 95 % confidence interval for sample size $T = 100$

The estimated autocorrelation coefficients converge to the true autocorrelation coefficient, in this case zero. The asymptotic distribution of $\sqrt{T}\hat{\rho}(h)$ converges to the standard normal distribution. This implies that for large T we can approximate the distribution of $\hat{\rho}(h)$ by a normal distribution with mean zero and variance $1/T$. This allows the construction of a 95 % confidence interval assuming that the true process is white noise. This confidence interval is therefore given by $\pm 1.96T^{-\frac{1}{2}}$. It can be used to verify if the observed process is indeed white noise.

Figure 4.1 plots the empirical autocorrelation function of a WN(0,1) process with a sample size of $T = 100$. The implied 95 % confidence interval is therefore equal to ± 0.196 . As each estimated autocorrelation coefficient falls within the confidence interval, we can conclude that the observed times series may indeed represent a white noise process.

Instead of examining each correlation coefficient separately, we can test the joint hypothesis that all correlation coefficients up to order N are simultaneously equal to zero, i.e. $\rho(1) = \rho(2) = \dots = \rho(N) = 0, N = 1, 2, \dots$. As each $\sqrt{T}\hat{\rho}(h)$ has an asymptotic standard normal distribution and is asymptotically uncorrelated with $\sqrt{T}\hat{\rho}(k), h \neq k$, the sum of the squared estimated autocorrelation coefficients is χ^2 distributed with N degrees of freedom. This test statistic is called *Box-Pierce statistic*:

$$Q = T \sum_{h=1}^N \hat{\rho}^2(h) \sim \chi_N^2.$$

A refinement of this test statistic is given by the *Ljung-Box statistic*:

$$Q' = T(T + 2) \sum_{h=1}^N \frac{\hat{\rho}^2(h)}{T - h} \sim \chi_N^2. \tag{4.4}$$

This test statistic is also asymptotically χ^2 distributed with the same degree of freedom N . This statistic accounts for the fact that the estimates for high orders h are based on a smaller number of observations and are thus less precise and more noisy. The two test statistics are used in the usual way. The null hypothesis that all correlation coefficients are jointly equal to zero is rejected if Q , respectively Q' is larger than the critical value corresponding the χ^2_N distribution. The number of summands N is usually taken to be rather large, for a sample size of 150 in the range between 15 and 20. The two test are also referred to as *Portmanteau tests*.

Example: MA(q) Process: $X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$
with $Z_t \sim \text{IID}(0, \sigma^2)$

In this case the covariance matrix is determined as

$$w_{ii} = 1 + 2\rho(1)^2 + \dots + 2\rho(q)^2 \quad \text{for } i > q.$$

For $i, j > q$, w_{ij} is equal to zero. The 95 % confidence interval for the MA(1) process $X_t = Z_t - 0.8Z_{t-1}$ is therefore given for a sample size of $T = 200$ by $\pm 1.96T^{-\frac{1}{2}}[1 + 2\rho(1)^2]^{\frac{1}{2}} = \pm 0.1684$.

Figure 4.2 shows the estimated autocorrelation function of the above MA(1) process together with 95 % confidence interval based on a white noise process and a MA(1) process with $\theta = -0.8$. As the first order autocorrelation coefficient is

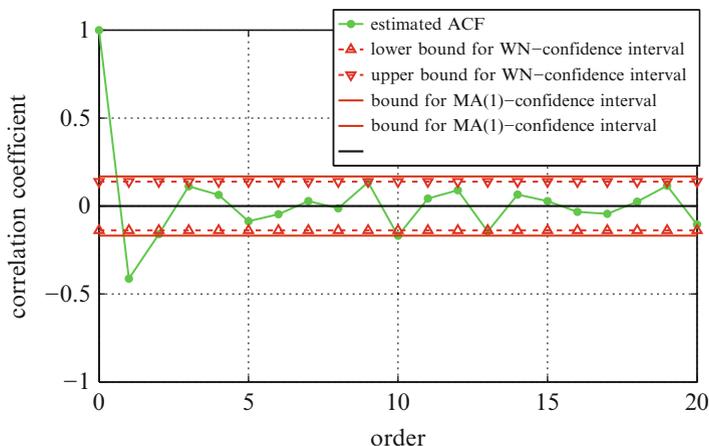


Fig. 4.2 Estimated autocorrelation function of a MA(1) process with $\theta = -0.8$ with corresponding 95 % confidence interval for $T = 200$

clearly outside the confidence interval whereas all other autocorrelation coefficients are inside it, the figure demonstrate that the observations are evidently the realization of MA(1) process.

Example: AR(1) Process $X_t - \phi X_{t-1} = Z_t$ with $Z_t \sim \text{IID}(0, \sigma^2)$

In this case the covariance matrix is determined as

$$\begin{aligned}
 w_{ii} &= \sum_{k=1}^i \phi^{2i} (\phi^k - \phi^{-k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k} (\phi^i - \phi^{-i})^2 \\
 &= \frac{(1 - \phi^{2i})(1 + \phi^2)}{1 - \phi^2} - 2i\phi^{2i} \\
 &\approx \frac{1 + \phi^2}{1 - \phi^2} \quad \text{for large } i.
 \end{aligned}$$

The formula for w_{ij} with $i \neq j$ are not shown. In any case, this formula is of relatively little importance because the partial autocorrelations are better suited for the identification of AR processes (see Sect. 3.5 and 4.3).

Figure 4.3 shows an estimated autocorrelation function of an AR(1) process. The autocorrelation coefficients decline exponentially which is a characteristic for an

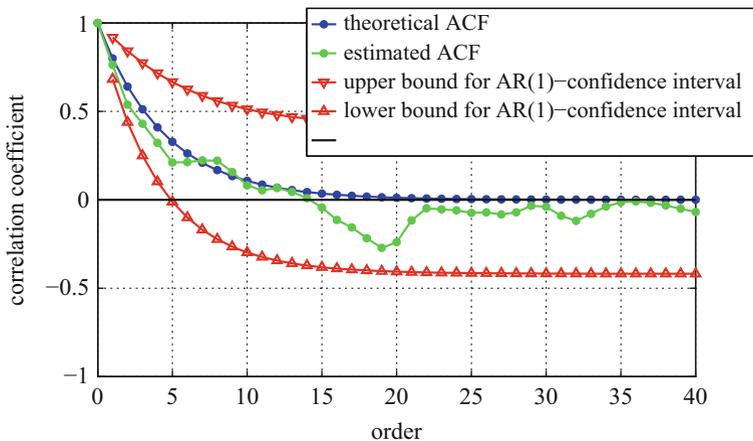


Fig. 4.3 Estimated autocorrelation function of an AR(1) process with $\phi = 0.8$ and corresponding 95 % confidence interval for $T = 100$

AR(1) process.⁴ Furthermore the coefficients are outside the confidence interval up to order 8 for white noise processes.

4.3 Estimation of the Partial Autocorrelation Function

According to its definition (see Definition 3.2), the partial autocorrelation of order h , $\alpha(h)$, is equal to a_h , the last element of the vector $\alpha_h = \Gamma_h^{-1} \gamma_h(1) = R_h^{-1} \rho_h(1)$. Thus, α_h and consequently a_h can be estimated by $\hat{\alpha}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h(1) = \hat{R}_h^{-1} \hat{\rho}_h(1)$. As $\rho(h)$ can be consistently estimated and is asymptotically normally distributed (see Sect. 4.2), the *continuous mapping theorem* (see Appendix C) ensures that the above estimator for $\alpha(h)$ is also consistent and asymptotically normal. In particular we have for an AR(p) process (Brockwell and Davis 1991)

$$\sqrt{T} \hat{\alpha}(h) \xrightarrow{d} N(0, 1) \quad \text{for } T \rightarrow \infty \text{ and } h > p.$$

This result allows to construct, as in the case of the autocorrelation coefficients, confidence intervals for the partial autocorrelations coefficients. The 95 % confidence interval is given by $\pm \frac{1.96}{\sqrt{T}}$. The AR(p) process is characterized by the fact that the partial autocorrelation coefficients are zero for $h > p$. $\hat{\alpha}(h)$ should therefore be inside the confidence interval for $h > p$ and outside for $h \leq p$. Figure 4.4 confirms this for an AR(1) process with $\phi = 0.8$.

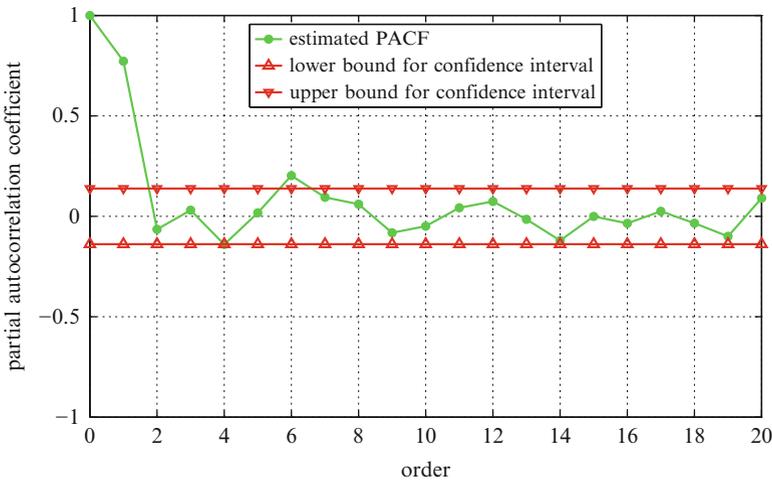


Fig. 4.4 Estimated PACF for an AR(1) process with $\phi = 0.8$ and corresponding 95 % confidence interval for $T = 200$

⁴As a reminder: the theoretical autocorrelation coefficients are $\rho(h) = \phi^{|h|}$.

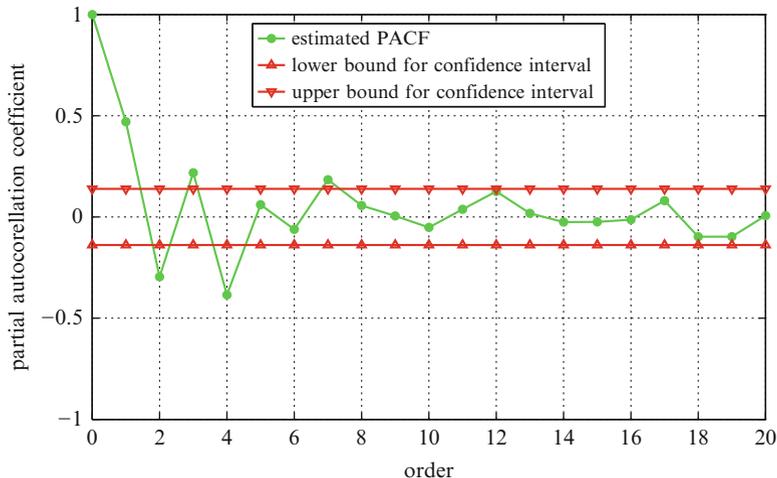


Fig. 4.5 Estimated PACF for a MA(1) process with $\theta = 0.8$ and corresponding 95 % confidence interval for $T = 200$

Figure 4.5 shows the estimated PACF for an MA(1) process with $\theta = 0.8$. In conformity with the theory, the partial autocorrelation coefficients converge to zero. They do so in an oscillating manner because θ is positive (see formula in Sect. 3.5).

4.4 Estimation of the Long-Run Variance

For many applications⁵ it is necessary to estimate the long-run variance J which is defined according to Eq. (4.1) as follows⁶

$$J = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) = \gamma(0) \left(1 + 2 \sum_{h=1}^{\infty} \rho(h) \right). \tag{4.5}$$

This can, in principle, be done in two different ways. The first one consists in the estimation of an ARMA model which is then used to derive the implied covariances as explained in Sect. 2.4. These covariances are then inserted into Eq. (4.5). The second method is a nonparametric one and is the subject for the rest of this Section. It has the advantage that it is not necessary to identify and estimate an appropriate ARMA model, a step which can be cumbersome in practice. Additional and more

⁵For example, when testing the null hypothesis $H_0: \mu = \mu_0$ in the case of serially correlated observations (see Sect. 4.1); for the Phillips-Perron unit-root test explained in Sect. 7.3.2.

⁶See Theorem 4.2 and the comments following it.

advanced material on this topic can be found in Andrews (1991), Andrews and Monahan (1992), or among others in Haan and Levin (1997).⁷

If the sample size is $T > 1$, only the covariances $\gamma(0), \dots, \gamma(T-1)$ can, in principle, be estimated. Thus, a first naive estimator of J is given by \hat{J}_T defined as

$$\hat{J}_T = \sum_{h=-T+1}^{T-1} \hat{\gamma}(h) = \hat{\gamma}(0) + 2 \sum_{h=1}^{T-1} \hat{\gamma}(h) = \hat{\gamma}(0) \left(1 + 2 \sum_{h=1}^{T-1} \hat{\rho}(h) \right),$$

where $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are the estimators for $\gamma(h)$ and $\rho(h)$, respectively, given in Sect. 4.2. As the estimators of the higher order autocovariances are based on smaller samples, their estimates become more erratic. At the same time, their weight in the above sum is the same as the lower order and more precisely estimated autocovariances. Thus, the higher order autocovariances have a disproportionate hazardous influence on the above estimator.

A remedy for this problem is to use only a certain number ℓ_T of autocovariances and/or to use a weighted sum instead of an unweighted one. This idea leads to the following class estimators:

$$\hat{J}_T = \hat{J}_T(\ell_T) = \frac{T}{T-r} \sum_{h=-T+1}^{T-1} k\left(\frac{h}{\ell_T}\right) \hat{\gamma}(h),$$

where k is a *weighting* or *kernel* function.⁸ The kernel functions are required to have the following properties:

- (i) $k : \mathbb{R} \rightarrow [-1, 1]$ is, with the exception of a finite number of points a continuous function. In particular, k is continuous at $x = 0$.
- (ii) k is quadratically integrable, i.e. $\int_{\mathbb{R}} k(x)^2 dx < \infty$;
- (iii) $k(0) = 1$;
- (iv) k is symmetric, i.e. $k(x) = k(-x)$ for all $x \in \mathbb{R}$.

The basic idea of the kernel function is to give relatively little weight to the higher order autocovariances and relatively more weight to the smaller order ones. As $k(0)$ equals one, the variance $\hat{\gamma}(0)$ receives weight one by construction. The continuity assumption implies that also the covariances of smaller order, i.e. for h small, receive a weight close to one. Table 4.1 lists some of the most popular kernel functions used in practice.

Figure 4.6 shows a plot of these functions. The first three functions are nonzero only for $|x| < 1$. This implies that only the orders h for which $|h| \leq \ell_T$ are taken into account. ℓ_T is called the *lag truncation parameter* or the *bandwidth*. The quadratic spectral kernel function is an example of a kernel function which takes all

⁷Note the connection between the long-run variance and the spectral density at frequency zero: $J = 2\pi f(0)$ where f is the spectral density function (see Sect. 6.3).

⁸Kernel functions are also relevant for spectral estimators. See in particular Sect. 6.3.

Table 4.1 Common kernel functions

Name	$k(x) =$
Boxcar (“truncated”)	1
Bartlett	$1 - x $
Daniell	$\frac{\sin(\pi x)}{\pi x}$
Tukey-Hanning	$(1 + \cos(\pi x))/2$
Quadratic Spectral	$\frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$

The function are, with the exception of the quadratic spectral function, only defined for $|x| \leq 1$. Outside this interval they are set to zero

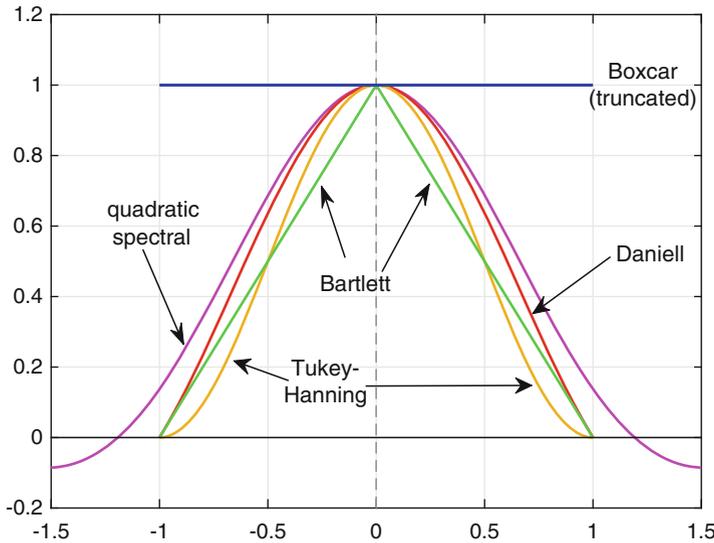


Fig. 4.6 Common kernel functions

covariances into account. Note that some weights are negative in this case as shown in Fig. 4.6.⁹

The estimator for the long-run variance is subject to the correction term $\frac{T}{T-r}$. This factor depends on the number of parameters estimated in a first step and is only relevant when the sample size is relatively small. In the case of the estimation of the mean r would be equal to one and the correction term is negligible. If on the other hand $X_t, t = 1, \dots, T$, are the residuals from multivariate regression, r designates the number of regressors. In many applications the correction term is omitted.

The lag truncation parameter or bandwidth, ℓ_T , depends on the number of observations. It is intuitive that the number of autocovariances accounted for in

⁹Phillips (2004) has proposed a nonparametric regression-based method which does not require a kernel function.

the computation of the long-run variance should increase with the sample size, i.e. we should have $\ell_T \rightarrow \infty$ for $T \rightarrow \infty$.¹⁰ The relevant issue is, at which rate the lag truncation parameter should go to infinity. The literature made several suggestions.¹¹ In the following we concentrate on the Bartlett and the quadratic spectral kernel because these function always deliver a positive long-run variance in small samples. Andrews (1991) proposes the following formula to determine the optimal bandwidth:

$$\text{Bartlett :} \quad \ell_T = 1.1447 [\alpha_{\text{Bartlett}} T]^{\frac{1}{3}}$$

$$\text{Quadratic Spectral :} \quad \ell_T = 1.3221 [\alpha_{\text{QuadraticSpectral}} T]^{\frac{1}{5}}$$

where $[\cdot]$ rounds to the nearest integer. The two coefficients α_{Bartlett} and $\alpha_{\text{QuadraticSpectral}}$ are data dependent constants which have to be determined in a first step from the data (see (Andrews 1991, 832–839), (Andrews and Monahan 1992, 958) and (Haan and Levin 1997)). If the underlying process is approximated by an AR(1) model, we get:

$$\alpha_{\text{Bartlett}} = \frac{4\hat{\rho}^2}{(1 - \hat{\rho}^2)(1 + \hat{\rho}^2)}$$

$$\alpha_{\text{QuadraticSpectral}} = \frac{4\hat{\rho}^2}{(1 - \hat{\rho})^4},$$

where $\hat{\rho}$ is the first order empirical autocorrelation coefficient.

In order to avoid the cumbersome determination of the α 's Newey and West (1994) suggest the following rules of thumb:

$$\text{Bartlett :} \quad \ell_T = \beta_{\text{Bartlett}} \left[\frac{T}{100} \right]^{\frac{2}{9}}$$

$$\text{Quadratic Spectral :} \quad \ell_T = \beta_{\text{QuadraticSpectral}} \left[\frac{T}{100} \right]^{\frac{2}{25}}.$$

It has been shown that values of 4 for β_{Bartlett} as well as for $\beta_{\text{QuadraticSpectral}}$ lead to acceptable results. A comparison of these formulas with the ones provided by Andrews shows that the latter imply larger values for ℓ_T when the sample sizes gets

¹⁰This is true even when the underlying process is known to be a MA(q) process. Even in this case it is advantageous to include also the autocovariances for $h > q$. The reason is twofold. First, only when $\ell_T \rightarrow \infty$ for $T \rightarrow \infty$, do we get a consistent estimator, i.e. $\hat{J}_T \rightarrow J_T$, respectively J . Second, the restriction to $\hat{\gamma}(h), |h| \leq q$, does not necessarily lead to positive value for the estimated long-run variance \hat{J}_T , even when the Bartlett kernel is used. See Ogaki (1992) for details.

¹¹See Haan and Levin (1997) for an overview.

larger. Both approaches lead to consistent estimates, i.e. $\hat{J}_T(\ell_T) - J_T \xrightarrow{P} 0$ for $T \rightarrow \infty$.

In practice, a combination of both parametric and nonparametric methods proved to deliver the best results. This combined method consists of five steps:

- (i) The first step is called *prewhitening* and consists in the estimation of a simple ARMA model for the process $\{X_t\}$ to remove the most obvious serial correlations. The idea, which goes back to Press and Tukey (1956) (see also Priestley (1981)), is to get a process for the residuals \hat{Z}_t which is close to a white noise process. Usually, an AR(1) model is sufficient.¹²
- (ii) Choose a kernel function and, if the method of Andrews has been chosen, the corresponding data dependent constants, i.e. α_{Bartlett} or $\alpha_{\text{QuadraticSpectral}}$ for the Bartlett, respectively the quadratic spectral kernel function.
- (iii) Compute the lag truncation parameter for the residuals using the above formulas.
- (iv) Estimate the long-run variance for the residuals \hat{Z}_t .
- (v) Compute the long-run variance for the original time series $\{X_t\}$.

If in the first step an AR(1) model, $X_t = \phi X_{t-1} + Z_t$, was used, the last step is given by:

$$\hat{J}_T^X(\ell_T) = \frac{\hat{J}_T^Z(\ell_T)}{(1 - \hat{\phi})^2},$$

where $\hat{J}_T^Z(\ell_T)$ and $\hat{J}_T^X(\ell_T)$ denote the estimated long-run variances of $\{X_t\}$ and $\{\hat{Z}_t\}$. In the general case, of an arbitrary ARMA model, $\Phi(L)X_t = \Theta(L)Z_t$, we get:

$$\hat{J}_T^X(\ell_T) = \left(\frac{\Theta(1)}{\Phi(1)} \right)^2 \hat{J}_T^Z(\ell_T).$$

4.4.1 An Example

Suppose we want to test whether the yearly growth rate of Switzerland's real GDP in the last 25 years was higher than 1 %. For this purpose we compute the percentage change against the corresponding quarter of the last year over the period 1982:1 to 2006:1 (97 observations in total), i.e. we compute $X_t = (1 - L^4) \log(\text{GDP}_t)$. The arithmetic average of these growth rates is 1.4960 with a variance of 3.0608.

¹²If in this step an AR(1) model is used and if a first order correlation $\hat{\phi}$ larger in absolute terms than 0.97 is obtained, Andrews and Monahan (1992, 457) suggest to replace $\hat{\phi}$ by -0.97 , respectively 0.97. Instead of using an arbitrary fixed value, it turns out that a data driven value is superior. Sul et al. (2005) suggest to replace 0.97 by $1 - 1/\sqrt{T}$ and -0.97 by $-1 + 1/\sqrt{T}$.

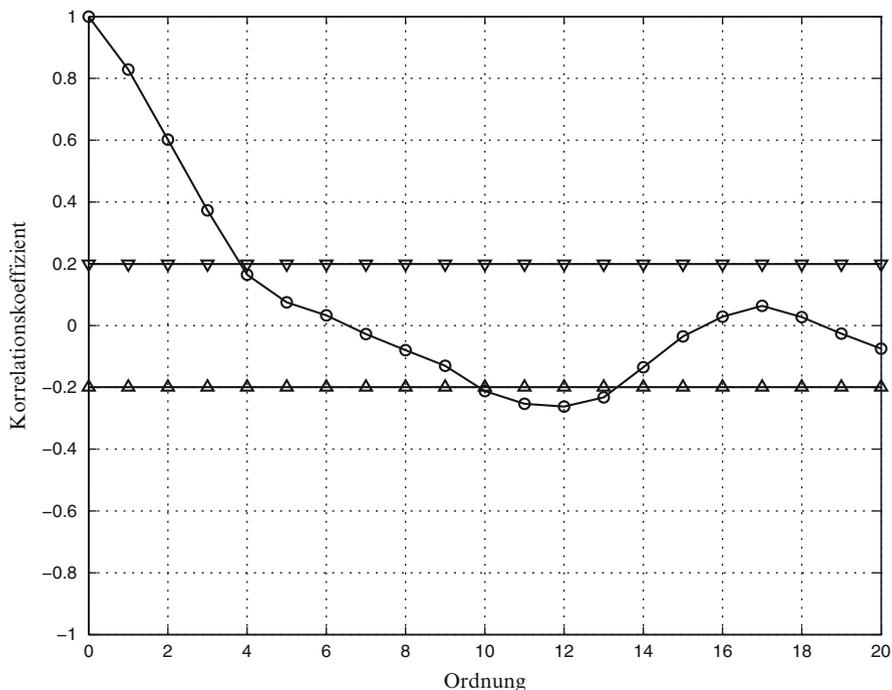


Fig. 4.7 Estimated autocorrelation function for Switzerland's real GDP growth (percentage change against corresponding last year's quarter)

We test the null hypothesis that the growth rate is smaller than one against the alternative that it is greater than one. The corresponding value of the t-statistic is $(1.4960 - 1) / \sqrt{3.0608/97} = 2.7922$. Taking a 5% significance level, the critical value for this one-sided test is 1.661. Thus the null hypothesis is clearly rejected.

The above computation is, however, not valid because the serial correlation of the time series was not taken into account. Indeed the estimated autocorrelation function shown in Fig. 4.7 clearly shows that the growth rate is subject to high and statistically significant autocorrelations.

Taking the Bartlett function as the kernel function, the rule of thumb formula for the lag truncation parameter suggest $\ell_T = 4$. The weights in the computation of the long-run variance are therefore

$$k(h/\ell_T) = \begin{cases} 1, & h = 0; \\ 3/4, & h = \pm 1; \\ 2/4, & h = \pm 2; \\ 1/4, & h = \pm 3; \\ 0, & |h| \geq 4. \end{cases}$$

The corresponding estimate for the long-run variance is therefore given by:

$$\hat{J}_T = 3.0608 \left(1 + 2\frac{3}{4}0.8287 + 2\frac{2}{4}0.6019 + 2\frac{1}{4}0.3727 \right) = 9.2783.$$

Using the long-run variance instead of the simple variance leads to a quite different value of the t-statistic: $(1.4960 - 1)/\sqrt{9.2783/97} = 1.6037$. The null hypothesis is thus not rejected at the 5% significance level when the serial correlation of the process is taken into account.

4.5 Exercises

Exercise 4.5.1. *You regress 100 realizations of a stationary stochastic process $\{X_t\}$ against a constant c . The least-squares estimate of c equals $\hat{c} = 0.04$ with an estimated standard deviation of $\hat{\sigma}_c = 0.15$. In addition, you have estimated the autocorrelation function up to order $h = 5$ and obtained the following values:*

$$\hat{\rho}(1) = -0.43, \hat{\rho}(2) = 0.13, \hat{\rho}(3) = -0.12, \hat{\rho}(4) = 0.18, \hat{\rho}(5) = -0.23.$$

- (i) *How do you interpret the estimated parameter value of 0.04?*
- (ii) *Examine the autocorrelation function. Do you think that $\{X_t\}$ is white noise?*
- (iii) *Why is the estimated standard deviation $\hat{\sigma}_c = 0.15$ incorrect?*
- (iv) *Estimate the long-run variance using the Bartlett kernel.*
- (v) *Test the null hypothesis that $\{X_t\}$ is a mean-zero process.*