Estimation of ARMA Models

5

The specification and estimation of an ARMA(p,q) model for a given realization involves several intermingled steps. First one must determine the orders p and q. Given the orders one can then estimate the parameters ϕ_j, θ_j and σ^2. Finally, the model has to pass several robustness checks in order to be accepted as a valid model. These checks may involve tests of parameter constancy, forecasting performance or tests for the inclusion of additional exogenous variables. This is usually an iterative process in which several models are examined. It is rarely the case that one model imposes itself. All too often, one is confronted in the modeling process with several trade-offs, like simple versus complex models or data fit versus forecasting performance. Finding the right balance among the different dimensions therefore requires some judgement based on experience.

We start the discussion by assuming that the orders of the ARMA process is known and the problem just consists in the estimation of the corresponding parameters from a realization of length T. For simplicity, we assume that the data are mean adjusted. We will introduce three estimation methods. The first one is a method of moments procedure where the theoretical moments are equated to the empirical ones. This procedure is known under the name of Yule-Walker estimator. The second procedure interprets the stochastic difference as a regression model and estimates the parameters by ordinary least-squares (OLS). These two methods work well if the underlying model is just an AR model and thus involves no MA terms. If the model comprises MA terms, a maximum likelihood (ML) approach must be pursued.

5.1 The Yule-Walker Estimator

We assume that the stochastic process has mean zero and is governed by a causal purely autoregressive model of order p:

$$\Phi(\mathrm{L})X_t = Z_t \qquad \text{with } Z_t \sim \mathrm{WN}(0, \sigma^2)$$

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_5

where $\Phi(\mathrm{L}) = 1-\phi_1\mathrm{L}-\phi_2\mathrm{L}^2-\ldots-\phi_p\mathrm{L}^p$. Causality with respect to $\{Z_t\}$ implies that there exists a sequence $\{\psi_j\}$ with $\sum_{j=0}^{\infty}|\psi_j| < \infty$ such that $X_t = \sum_{j=0}^{\infty}\psi_j Z_{t-j} = \Psi(\mathrm{L})Z_t$. Multiplying the above difference equation by $X_{t-j}, j = 0, 1, \ldots, p$ and taking expectations leads to the following equation system for the parameters $\Phi = (\phi_1, \ldots, \phi_p)'$ and σ^2:

$$\begin{aligned}\gamma(0) - \phi_1\gamma(1) - \ldots - \phi_p\gamma(p) &= \sigma^2\\ \gamma(1) - \phi_1\gamma(0) - \ldots - \phi_p\gamma(p-1) &= 0\\ &\ldots\\ \gamma(p) - \phi_1\gamma(p-1) - \ldots - \phi_p\gamma(0) &= 0\end{aligned}$$

This equation system is known as the *Yule-Walker equations*. It can be written compactly in matrix algebra as:

$$\gamma(0) - \Phi'\gamma_p(1) = \sigma^2,$$

$$\begin{pmatrix}\gamma(0) & \gamma(1) & \ldots & \gamma(p-1)\\ \gamma(1) & \gamma(0) & \ldots & \gamma(p-2)\\ \vdots & \vdots & \ddots & \vdots\\ \gamma(p-1) & \gamma(p-2) & \ldots & \gamma(0)\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_p\end{pmatrix} = \begin{pmatrix}\gamma(1)\\ \gamma(2)\\ \vdots\\ \gamma(p)\end{pmatrix},$$

respectively

$$\begin{aligned}\gamma(0) - \Phi'\gamma_p(1) &= \sigma^2,\\ \Gamma_p\Phi &= \gamma_p(1).\end{aligned}$$

The *Yule-Walker estimator* is obtained by replacing the theoretical moments by the empirical ones and solving the resulting equation system for the unknown parameters:

$$\begin{aligned}\widehat{\Phi} &= \widehat{\Gamma}_p^{-1}\,\hat{\gamma}_p(1) = \widehat{R}_p^{-1}\,\hat{\rho}_p(1)\\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \widehat{\Phi}'\,\hat{\gamma}_p(1) = \hat{\gamma}(0)\left(1 - \hat{\rho}_p(1)'\,\widehat{R}_p^{-1}\,\hat{\rho}_p(1)\right)\end{aligned}$$

Note the recursiveness of the equation system: the estimate $\widehat{\Phi}$ is obtained without knowledge of $\hat{\sigma}^2$ as the estimator $\widehat{R}_p^{-1}\,\hat{\rho}_p(1)$ involves only autocorrelations. The estimates $\widehat{\Gamma}_p$, $\widehat{R}_p$, $\hat{\gamma}_p(1)$, $\hat{\rho}_p(1)$, and $\hat{\gamma}(0)$ are obtained in the usual way as explained in Chap. 4.[1]

[1]Note that the application of the estimator introduced in Sect. 4.2 guarantees that $\widehat{\Gamma}_p$ is always invertible.

The construction of the Yule-Walker estimator implies that the first p values of the autocovariance, respectively the autocorrelation function, implied by the estimated model exactly correspond to their estimated counterparts. It can be shown that this moment estimator always delivers coefficients $\widehat{\Phi}$ which imply that $\{X_t\}$ is causal with respect to $\{Z_t\}$. In addition, the following Theorem establishes that the estimated coefficients are asymptotically normal.

Theorem 5.1 (Asymptotic Normality of Yule-Walker Estimator)**.** *Let $\{X_t\}$ be an AR(p) process which is causal with respect to $\{Z_t\}$ whereby $\{Z_t\} \sim \text{IID}(0, \sigma^2)$. Then the Yule-Walker estimator is consistent and $\widehat{\Phi}$ is asymptotically normal with distribution given by:*

$$\sqrt{T}\left(\widehat{\Phi} - \Phi\right) \xrightarrow{d} \text{N}\left(0, \sigma^2 \Gamma_p^{-1}\right).$$

In addition we have that

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

Proof. See Brockwell and Davis (1991, 233–234). □

Noteworthy, the asymptotic covariance matrix of the Yule-Walker estimate is independent of σ^2. In practice, the unknown parameters $\sigma^2 \Gamma_p^{-1}$ are replaced by their empirical counterparts.

Example: AR(1) Process

In the case of an AR(1) process, the Yule-Walker equation is $\widehat{\Gamma}_1 \Phi = \hat{\gamma}_1(0)$ which simplifies to $\hat{\gamma}(0)\phi = \hat{\gamma}(1)$. The Yule-Walker estimator thus becomes:

$$\widehat{\Phi} = \hat{\phi} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \hat{\rho}(1).$$

The asymptotic distribution then is

$$\sqrt{T}\left(\hat{\phi} - \phi\right) \xrightarrow{\text{d}} \text{N}\left(0, \frac{\sigma^2}{\gamma(0)}\right) = \text{N}\left(0, 1 - \phi^2\right).$$

This shows that the assumption of causality, i.e. $|\phi| < 1$, is crucial. Otherwise no strictly positive value for the variance would exist. For the case $\phi = 1$ which corresponds to the random walk, the asymptotic distribution of $\sqrt{T}(\hat{\phi} - 1)$ becomes degenerate as the variance is equal to zero. This case is, however, of prime importance in economics and is treated detail in Chap. 7.

In practice the order of the model is usually unknown. However, one can expect when estimating an AR(m) model whereby the true order p is strictly smaller than m

that the estimated coefficients $\hat{\phi}_{p+1}, \ldots, \hat{\phi}_m$ should be close to zero. This is indeed the case as shown in Brockwell and Davis (1991, 241). In particular, under the assumptions of Theorem 5.1 it holds that

$$\sqrt{T}\,\hat{\phi}_m \xrightarrow{\text{d}} \text{N}(0,1) \qquad \text{for } m > p. \tag{5.1}$$

This result justifies the following strategy to identify the order of an AR-model. Estimate in a first step a highly parameterized model (overfitted model), i.e. a model with a large value of m, and test via a t-test whether ϕ_m is zero. If the hypothesis cannot be rejected, reduce the order of the model from m to $m-1$ and repeat the same procedure now with respect to ϕ_{m-1}. This is done until the hypothesis can no longer be rejected.

If the order of the initial model is too low (underfitted model) so that the true order is higher than m, one incurs an "omitted variable bias". The corresponding estimates are no longer consistent. In Sect. 5.4, we take closer look at the problem of determining the order of a model.

Example: MA(q) Process

The Yule-Walker estimator can, in principle, also be applied to MA(q) or ARMA(p,q) processes with $q > 0$. However, the analysis of the simple MA(1) process in Sect. 1.5.1 showed that the relation between the autocorrelations and the model parameters is nonlinear and may have two, one, or no solution. Consider again the MA(1) process as an example. It is given by the stochastic difference equation $X_t = Z_t + \theta Z_{t-1}$ with $Z_t \sim \text{IID}(0, \sigma^2)$. The Yule-Walker equations are then as follows:

$$\hat{\gamma}(0) = \hat{\sigma}^2(1 + \hat{\theta}^2)$$
$$\hat{\gamma}(1) = \hat{\sigma}^2\hat{\theta}$$

As shown in Sect. 1.5.1, this system of equations has for the case $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| < 1/2$ two solutions; for $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| = 1/2$ one solution; and for $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| > 1/2$ no real solution. In the case of several solutions, we usually take the invertible one which leads to $|\theta| < 1$. Invertibility is, however, a restriction which is hard to implement in the case of higher order MA processes. Moreover, it can be shown that Yule-Walker estimator is no longer consistent in general (see Brockwell and Davis (1991, 246) for details). For these reasons, it is not advisable to use the Yule-Walker estimator in the case of MA processes, especially when there exist consistent and efficient alternatives.

5.2 Ordinary Least-Squares (OLS) Estimation of an AR(p) Model

An alternative approach is to view the AR model as a regression model for X_t with regressors $X_{t-1}, \ldots, X_{t-p}$ and error term Z_t:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t, \qquad Z_t \sim \mathrm{WN}(0, \sigma^2).$$

Given observation for $X_1, \ldots, X_T$, the regression model can be compactly written in matrix algebra as follows:

$$\begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} X_p & X_{p-1} & \ldots & X_1 \\ X_{p+1} & X_p & \ldots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{T-1} & X_{T-2} & \ldots & X_{T-p} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_T \end{pmatrix},$$

$$Y = \mathbf{X}\Phi + Z. \tag{5.2}$$

Note that the first p observations are lost and that the effective sample size is thus reduced to $T - p$. The least-squares estimator (OLS estimator) is obtained as the minimizer of the sum of squares $S(\Phi)$:

$$\begin{aligned} S(\Phi) = Z'Z &= (Y - \mathbf{X}\Phi)'(Y - \mathbf{X}\Phi) \\ &= \sum_{t=p+1}^{T} \left(X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p}\right)^2 \\ &= \sum_{t=p+1}^{T} \left(X_t - \mathbb{P}_{t-1} X_t\right)^2 \longrightarrow \min_{\Phi}. \end{aligned} \tag{5.3}$$

Note that the optimization problem involves no constraints, in particular causality is not imposed as a restriction. The solution of this minimization problem is given by usual formula:

$$\widehat{\Phi} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\mathbf{X}'Y\right).$$

Though Eq. (5.2) resembles very much an ordinary regression model, there are some important differences. First, the standard orthogonality assumption between regressors and error is violated. The regressors $X_{t-j}, j = 1, \ldots, p$, are correlated with the error terms $Z_{t-j}, j = 1, 2, \ldots$. Second, there is a dependency on the starting values $X_p, \ldots, X_1$. The assumption of causality, however, insures that these features do not play a role asymptotically. It can be shown that $(\mathbf{X}'\mathbf{X})/T$ converges in probability to $\widehat{\Gamma}_p$ and $(\mathbf{X}'Y)/T$ to $\hat{\gamma}_p$. In addition, under quite general conditions,

$T^{-1/2}\mathbf{X}'Z$ is asymptotically normally distributed with mean 0 and variance $\sigma^2\Gamma_p$. Then by Slutzky's Lemma C.10, $\sqrt{T}(\widehat{\Phi} - \Phi) = \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{X}'Z}{\sqrt{T}}\right)$ converges in distribution to $\mathrm{N}(0, \sigma^2\Gamma_p^{-1})$ Thus, the OLS estimator is asymptotically equivalent to the Yule-Walker estimator.

Theorem 5.2 (Asymptotic Normality of the Least-Squares Estimator)**.** *Under the same assumptions as in Theorem 5.1, the ordinary least-squares estimator (OLS estimator)* $\widehat{\Phi} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'Y)$ *is asymptotically distributed as*

$$\sqrt{T}\left(\widehat{\Phi} - \Phi\right) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2\Gamma_p^{-1}\right),$$

$$\operatorname{plim} s_T^2 = \sigma^2$$

where $s_T^2 = \widehat{Z}'\widehat{Z}/T$ *and* $\widehat{Z}_t$ *are the OLS residuals.*

Proof. See Chap. 13 and in particular Sect. 13.3 for a proof in the multivariate case. Additional details may be gathered from Brockwell and Davis (1991, chapter 8). □

Remark 5.1. In practice $\sigma^2\Gamma_p^{-1}$ is approximated by $s_T^2(\mathbf{X}'\mathbf{X}/T)^{-1}$. Thus, for large T, $\widehat{\Phi}$ can be viewed as being normally distributed as $\mathrm{N}(\Phi, s_T^2(\mathbf{X}'\mathbf{X})^{-1})$. This result allows the application of the usual t- and F-tests.

Because the regressors X_{t-j}, $j = 1, \ldots, p$ are correlated with the errors terms Z_{t-j}, $j = 1, 2, \ldots$, the Gauss-Markov theorem cannot be applied. This implies that the least-squares estimator is no longer unbiased in finite samples. It can be shown that the estimates of an AR(1) model are downward biased when the true value of ϕ is between zero and one. MacKinnon and Smith (1998, figure 1) plots the bias as a function of the sample size and the true parameter (see also Fig. 7.1). As the bias function is almost linear in the range $-0.85 < \phi < 0.85$, an approximately unbiased estimator for the AR(1) model has been proposed by Marriott and Pope (1954), Kendall (1954), and Orcutt and Winokur (1969) (for further details see MacKinnon and Smith 1998):

$$\hat{\phi}_{\text{corrected}} = \frac{1}{T-3}(T\hat{\phi}_{\text{OLS}} + 1).$$

Remark 5.2. The OLS estimator does in general not deliver coefficients $\hat{\Phi}$ for which $\{X_t\}$ is causal with respect $\{Z_t\}$. In particular, in the case of an AR(1) model, it can happen that, in contrast to the Yule-Walker estimator, $|\hat{\phi}|$ is larger than one despite the fact that the true parameter is absolutely smaller than one. Nevertheless, the least-squares estimator is to be preferred in practice because it delivers small-sample biases of the coefficients which are smaller than those of Yule-Walker estimator, especially for roots of $\Phi(z)$ close to the unit circle (Tjøstheim and Paulsen 1983; Shaman and Stine 1988; Reinsel 1993).

Appendix: Proof of the Asymptotic Normality of the OLS Estimator

The proofs of Theorems 5.1 and 5.2 are rather involved and will therefore not be pursued here. A proof for the more general multivariate case will be given in Chap. 13. It is, however, instructive to look at a simple case, namely the AR(1) model with $|\phi| < 1, Z_t \sim \text{IIN}(0, \sigma^2)$ and $X_0 = 0$. Denoting by $\hat{\phi}_T$ the OLS estimator of ϕ, we have:

$$\sqrt{T}\left(\hat{\phi}_T - \phi\right) = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}Z_t}{\frac{1}{T}\sum_{t=1}^{T} X_{t-1}^2}. \tag{5.4}$$

Moreover, X_t can be written as follows:

$$X_t = Z_t + \phi Z_{t-1} + \ldots + \phi^{t-1} Z_1.$$

By assumption each Z_j, $j = 1, \ldots, t$, is normally distributed so that X_t as a sum normally distributed random variables is also normally distributed. Because the Z_j's are independent we have: $X_t \sim \text{N}\left(0, \sigma^2 \frac{1-\phi^{2t}}{1-\phi^2}\right)$.

The expected value of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}Z_t$ is zero because $Z_t \sim \text{IIN}(0, \sigma^2)$. The variance of this expression is given by

$$\begin{aligned}\mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}Z_t\right) &= \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}X_{t-1}^2 Z_t^2 + \frac{2}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1} \mathbb{E}Z_t \mathbb{E}X_{t-1}X_j Z_j \\ &= \frac{\sigma^2}{T}\sum_{t=1}^{T} \mathbb{E}X_{t-1}^2.\end{aligned}$$

Moreover, $\sum_{t=1}^{T} X_t^2 = \sum_{t=1}^{T} X_{t-1}^2 - (X_0^2 - X_T^2) = \phi^2 \sum_{t=1}^{T} X_{t-1}^2 + \sum_{t=1}^{T} Z_t^2 + 2\phi \sum_{t=1}^{T} X_{t-1}Z_t$ so that

$$\sum_{t=1}^{T} X_{t-1}^2 = \frac{1}{1-\phi^2}(X_0^2 - X_T^2) + \frac{1}{1-\phi^2}\sum_{t=1}^{T} Z_t^2 + \frac{2\phi}{1-\phi^2}\sum_{t=1}^{T} X_{t-1}Z_t.$$

The expected value multiplied by σ^2/T thus is equal to

$$\begin{aligned}\frac{\sigma^2}{T}\sum_{t=1}^{T} \mathbb{E}X_{t-1}^2 &= \frac{\sigma^2}{1-\phi^2}\frac{\mathbb{E}X_0^2 - \mathbb{E}X_T^2}{T} \\ &\quad + \frac{\sigma^2}{1-\phi^2}\frac{\sum_{t=1}^{T} \mathbb{E}Z_t^2}{T} + \frac{2\phi}{1-\phi^2}\frac{\sum_{t=1}^{T} X_{t-1}Z_t}{T} \\ &= -\frac{\sigma^4(1-\phi^{2T})}{T(1-\phi^2)^2} + \frac{\sigma^4}{1-\phi^2}.\end{aligned}$$

For T going to infinity, we finally get:

$$\lim_{T\to\infty} \mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}Z_t\right) = \frac{\sigma^4}{1-\phi^2}.$$

The numerator in Eq. (5.4) therefore converges to a normal random variable with mean zero and variance $\frac{\sigma^4}{1-\phi^2}$.

The denominator in Eq. (5.4) can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T} X_{t-1}^2 = \frac{X_0^2 - X_T^2}{(1-\phi^2)T} + \frac{1}{(1-\phi^2)T}\sum_{t=1}^{T} Z_t^2 + \frac{2\phi}{(1-\phi^2)T}\sum_{t=1}^{T} X_{t-1}Z_t.$$

The expected value and the variance of X_T^2/T converge to zero. Chebyschev's inequality (see Theorem C.3 in Appendix C) then implies that the first term converges also in probability to zero. X_0 is equal to zero by assumption. The second term has a constant mean equal to $\sigma^2/(1-\phi^2)$ and a variance which converges to zero. Theorem C.8 in Appendix C then implies that the second term converges in probability to $\sigma^2/(1-\phi^2)$. The third term has a mean of zero and a variance which converges to zero. Thus the third term converges to zero in probability. This implies:

$$\frac{1}{T}\sum_{t=1}^{T} X_{t-1}^2 \xrightarrow{\text{p}} \frac{\sigma^2}{1-\phi^2}.$$

Putting the results for the numerator and the denominator together and applying Theorem C.10 and the continuous mapping theorem for the convergence in distribution one finally obtains:

$$\sqrt{T}\left(\hat{\phi}_T - \phi\right) \xrightarrow{\text{d}} \text{N}(0, 1-\phi^2). \tag{5.5}$$

Thereby the value for the variance is derived from

$$\frac{\sigma^4}{1-\phi^2} \times \frac{1}{\left(\frac{\sigma^2}{1-\phi^2}\right)^2} = 1-\phi^2.$$

5.3 Estimation of an ARMA(p,q) Model

While the estimation of AR models by OLS is rather straightforward and leads to consistent and asymptotically efficient estimates, the estimation of ARMA models is more complex. The reason is that, in contrast to past X_t's $Z_t, Z_{t-1}, \ldots, Z_{t-q}$ are not directly observable from the data. They must be inferred from the observations of X_t. The standard method for the estimation of ARMA models is the method of maximum likelihood which will be explained in this section.

We assume that the process $\{X_t\}$ is a causal and invertible ARMA(p,q) process following the difference equation

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

with $Z_t \sim \text{IID}(0, \sigma^2)$. We also assume that $\Phi(z)$ and $\Theta(z)$ have no roots in common. We then stack the parameters of the model into a vector β and a scalar σ^2:

$$\beta = \left(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q\right)' \quad \text{and} \quad \sigma^2.$$

Given the assumption above the admissible parameter space for β, $\mathcal{C}$, is described by the following set:

$$\begin{aligned}\mathcal{C} = \{\beta \in \mathbb{R}^{p+q} : \Phi(z)\Theta(z) \neq 0 \text{ for } |z| \leq 1, \phi_p\theta_q \neq 0,\\ \Phi(z) \text{ and } \Theta(z) \text{ have no roots in common}\}\end{aligned}$$

The estimation by the *method of maximum likelihood* (ML method) is based on some assumption about the joint distribution of $\mathbf{X}_T = (X_1, \ldots, X_T)'$ given the parameters β and σ^2. This joint distribution function is called the likelihood function. The method of maximum likelihood then determines the parameters such that the probability of observing a given sample $\mathbf{x}_T = (x_1, \ldots, x_T)$ is maximized. This is achieved by maximizing the likelihood function with respect to the parameters.

By far the most important case is given by assuming that $\{X_t\}$ is a Gaussian process with mean zero and autocovariance function γ. This implies that $\mathbf{X}_T = (X_1, \ldots, X_T)'$ is distributed as a multivariate normal with mean zero and variance Γ_T.[2] The *Gaussian likelihood function* given the observations $\mathbf{x}_T$, $\text{L}_T(\beta, \sigma^2|\mathbf{x}_T)$, is then given by

$$\begin{aligned}\text{L}_T(\beta, \sigma^2|\mathbf{x}_T) &= (2\pi)^{-T/2}(\det \Gamma_T)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}_T'\Gamma_T^{-1}\mathbf{x}_T\right)\\ &= (2\pi\sigma^2)^{-T/2}(\det G_T)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\mathbf{x}_T'G_T^{-1}\mathbf{x}_T\right)\end{aligned}$$

where $G_T = \sigma^{-2}\Gamma_T$. Note that, in contrast to Γ_T, G_T does only depend on β and not on σ^2.[3] If one wants to point out the dependence of G_T from β we write $G_T(\beta)$. The method of maximum likelihood then consists in the maximization of the likelihood function with respect to β and σ^2 taking the data $\mathbf{x}_T$ as given.

[2] If the process does not have a mean of zero, we can demean the data in a preliminary step.

[3] In Sect. 2.4 we showed how the autocovariance function γ and as a consequence Γ_T, respectively G_T can be inferred from a given ARMA model, i.e from a given β.

The first order condition of this maximization problem with respect to σ^2 is obtained by taking the logarithm of the likelihood function $\mathrm{L}_T(\beta, \sigma^2|\mathbf{x}_T)$ and differentiating with respect σ^2 and setting the resulting equation equal to zero:

$$\frac{\partial \ln \mathrm{L}_T(\beta, \sigma^2|\mathbf{x}_T)}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{\mathbf{X}_T' G_T^{-1} \mathbf{X}_T}{2\sigma^4} = 0.$$

Solving this equation with respect to σ^2 we get as the solution: $\sigma^2 = T^{-1}\mathbf{x}_T' G_T^{-1}\mathbf{x}_T$. Inserting this value into the original likelihood function and taking the logarithm, one gets the concentrated log-likelihood function:

$$\ln \mathrm{L}_T(\beta|\mathbf{x}_T) = -\ln(2\pi) - \frac{T}{2}\ln\left(T^{-1}\mathbf{x}_T' G_T(\beta)^{-1}\mathbf{x}_T\right) - \frac{1}{2}\ln\det G_T(\beta) - \frac{T}{2}.$$

This function is then maximized with respect to $\beta \in \mathcal{C}$. This is, however, equivalent to minimizing the function

$$\ell_T(\beta|\mathbf{x}_T) = \ln\left(T^{-1}\mathbf{x}_T' G_T(\beta)^{-1}\mathbf{x}_T\right) + T^{-1}\ln\det G_T(\beta) \longrightarrow \min_{\beta\in\mathcal{C}}.$$

The value of β which minimizes the above function is called *maximum-likelihood estimator* of β. It will be denoted by $\hat{\beta}_{\mathrm{ML}}$. The maximum-likelihood estimator for σ^2, $\hat{\sigma}^2_{\mathrm{ML}}$, is then given by

$$\hat{\sigma}^2_{\mathrm{ML}} = T^{-1}\mathbf{x}_T' G_T(\hat{\beta}_{\mathrm{ML}})^{-1}\mathbf{x}_T.$$

The actual computation of $\det G_T(\beta)$ and $G_T(\beta)^{-1}$ is numerically involved, especially when T is large, and should therefore be avoided. It is therefore convenient to rewrite the likelihood function in a different, but equivalent form:

$$\begin{aligned}\mathrm{L}_T(\beta, \sigma^2|\mathbf{x}_T) =& (2\pi\sigma^2)^{-T/2}(r_0 r_1 \dots r_{T-1})^{-1/2} \\ & \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\frac{(X_t - \mathbb{P}_{t-1}X_t)^2}{r_{t-1}}\right).\end{aligned}$$

Thereby $\mathbb{P}_{t-1}X_t$ denotes least-squares predictor of X_t given $X_{t-1}, \dots, X_1$ and $r_t = v_t/\sigma^2$ where v_t is the mean squared forecast error as defined in Sect. 3.1. Several numerical algorithms have been developed to compute these forecast in a numerically efficient and stable way.[4]

$\mathbb{P}_{t-1}X_t$ and r_t do not depend on σ^2 so that the partial differentiation of the log-likelihood function $\ln \mathrm{L}(\beta, \sigma^2|\mathbf{x}_T)$ with respect to the parameters leads to the maximum likelihood estimator. This estimator fulfills the following equations:

[4] One such algorithm is the innovation algorithm. See Brockwell and Davis (1991, section 5) for details.

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{T} S(\hat{\beta}_{\mathrm{ML}}),$$

where

$$S(\hat{\beta}_{\mathrm{ML}}) = \sum_{t=1}^{T} \frac{(X_t - \mathbb{P}_{t-1} X_t)^2}{r_{t-1}}$$

and where $\hat{\beta}_{\mathrm{ML}}$ denote the value of β which minimizes the function

$$\ell_T(\beta|\mathbf{x}_T) = \ln\left(\frac{1}{T} S(\beta)\right) + \frac{1}{T}\sum_{t=1}^{T} \ln r_{t-1}$$

subject to $\beta \in \mathcal{C}$. This optimization problem must be solved numerically. In practice, one chooses as a starting value β_0 for the iteration an initial estimate such that $\beta_0 \in \mathcal{C}$. In the following iterations this restriction is no longer imposed to enhance speed and reduce the complexity of the optimization problem. This implies that one must check whether the so obtained final estimates are indeed in $\mathcal{C}$.

If instead of $\ell_T(\beta|\mathbf{x}_T)$, the function

$$S(\beta) = \sum_{t=1}^{T} \frac{(X_t - \mathbb{P}_{t-1} X_t)^2}{r_{t-1}}$$

is minimized subject to constraint $\beta \in \mathcal{C}$, we obtain the *least-squares estimator* of β denoted by $\hat{\beta}_{\mathrm{LS}}$. The least-squares estimator of σ^2, $\hat{\sigma}^2_{\mathrm{LS}}$, is then

$$\hat{\sigma}^2_{\mathrm{LS}} = \frac{S(\hat{\beta}_{\mathrm{LS}})}{T - p - q}.$$

The term $\frac{1}{T}\sum_{t=1}^{T} \ln r_{t-1}$ disappears asymptotically because, given the restriction $\beta \in \mathcal{C}$, the mean-squared forecast error v_T converges to σ^2 and thus r_T goes to one as T goes to infinity. This implies that for T going to infinity the maximization of the likelihood function becomes equivalent to the minimization of the least-squares criterion. Thus the maximum-likelihood estimator and the least-squares estimator share the same asymptotic normal distribution.

Note also that in the case of autoregressive models r_t is constant and equal to one. In this case, the least-squares criterion $S(\beta)$ reduces to the criterion (5.3) discussed in the previous Sect. 5.2.

Theorem 5.3 (Asymptotic Distribution of ML Estimator)**.** *If $\{X_t\}$ is an ARMA process with true parameters $\beta \in \mathcal{C}$ and $Z_t \sim \text{IID}(0, \sigma^2)$ with $\sigma^2 > 0$ then the maximum-likelihood estimator and the least-squares estimator have asymptotically the same normal distribution:*

$$\sqrt{T}\left(\hat{\beta}_{\text{ML}} - \beta\right) \xrightarrow{d} \text{N}\left(0, V(\beta)\right),$$

$$\sqrt{T}\left(\hat{\beta}_{\text{LS}} - \beta\right) \xrightarrow{d} \text{N}\left(0, V(\beta)\right).$$

The asymptotic covariance matrix $V(\beta)$ is thereby given by

$$V(\beta) = \begin{pmatrix} \mathbb{E}U_t U_t' & \mathbb{E}U_t V_t' \\ \mathbb{E}V_t U_t' & \mathbb{E}V_t V_t' \end{pmatrix}^{-1}$$

$$U_t = \left(u_t, u_{t-1}, \ldots, u_{t-p+1}\right)'$$

$$V_t = \left(v_t, v_{t-1}, \ldots, v_{t-q+1}\right)'$$

where $\{u_t\}$ and $\{v_t\}$ denote autoregressive processes defined as $\Phi(\text{L})u_t = w_t$ and $\Theta(\text{L})v_t = w_t$ with $w_t \sim \text{WN}(0, 1)$.

Proof. See Brockwell and Davis (1991, Section 8.8). □

It can be shown that both estimators are asymptotically efficient.[5] Note that the asymptotic covariance matrix $V(\beta)$ is independent of σ^2.

The use of the Gaussian likelihood function makes sense even when the process is not Gaussian. First, the Gaussian likelihood can still be interpreted as a measure of fit of the ARMA model to the data. Second, the asymptotic distribution is still Gaussian even when the process is not Gaussian as long as $Z_t \sim \text{IID}(0, \sigma^2)$. The Gaussian likelihood is then called the quasi Gaussian likelihood. The use of the Gaussian likelihood under this circumstance is, however, in general no longer efficient.

Example: AR(p) Process

In this case $\beta = (\phi_1, \ldots, \phi_p)$ and $V(\beta) = \left(\mathbb{E}U_t U_t'\right)^{-1} = \sigma^2 \Gamma_p^{-1}$. This is, however, the same asymptotic distribution as the Yule-Walker estimator. The Yule-Walker, the least-squares, and the maximum likelihood estimator are therefore asymptotically equivalent in the case of an AR(p) process. The main difference lies in the treatment of the first p observations.

[5] See Brockwell and Davis (1991) and Fan and Yao (2003) for details.

In particular, we have:

$$\begin{aligned}
&\text{AR(1)}: && \hat{\phi} \sim \mathrm{N}\left(\phi, (1-\phi^2)/T\right), \\
&\text{AR(2)}: && \begin{pmatrix}\hat{\phi}_1 \\ \hat{\phi}_2\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\phi_1 \\ \phi_2\end{pmatrix}, \frac{1}{T}\begin{pmatrix}1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2\end{pmatrix}\right).
\end{aligned}$$

Example: MA(q) Process

Similarly, one can compute the asymptotic distribution for an MA(q) process. In particular, we have:

$$\begin{aligned}
&\text{MA(1)}: && \hat{\theta} \sim \mathrm{N}\left(\theta, (1-\theta^2)/T\right), \\
&\text{MA(2)}: && \begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}, \frac{1}{T}\begin{pmatrix}1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2\end{pmatrix}\right).
\end{aligned}$$

Example: ARMA(1,1) Process

For an ARMA(1,1) process the asymptotic covariance matrix is given by

$$V(\phi, \theta) = \begin{pmatrix}\left(1-\phi^2\right)^{-1} & (1+\phi\theta)^{-1} \\ (1+\phi\theta)^{-1} & \left(1-\theta^2\right)^{-1}\end{pmatrix}^{-1}.$$

Therefore we have:

$$\begin{pmatrix}\hat{\phi} \\ \hat{\theta}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\phi \\ \theta\end{pmatrix}, \frac{1}{T}\,\frac{1+\phi\theta}{(\phi+\theta)^2}\begin{pmatrix}(1-\phi^2)(1+\phi\theta) & -(1-\theta^2)(1-\phi^2) \\ -(1-\theta^2)(1-\phi^2) & (1-\theta^2)(1+\phi\theta)\end{pmatrix}\right).$$

5.4 Estimation of the Orders p and q

Up to now we have always assumed that the true orders of the ARMA model p and q are known. This is, however, seldom the case in practice. As economic theory does usually not provide an indication, it is all too often the case that the orders of the ARMA model must be identified from the data. In such a situation one can make two type of errors: p and q are too large in which case we speak of overfitting; p and q are too low in which case we speak of underfitting.

In the case of overfitting, the maximum likelihood estimator is no longer consistent for the true parameter, but still consistent for the coefficients of the causal representation $\psi_j, j = 0, 1, 2, \ldots$, where $\psi(z) = \frac{\theta(z)}{\phi(z)}$. This can be illustrated by the

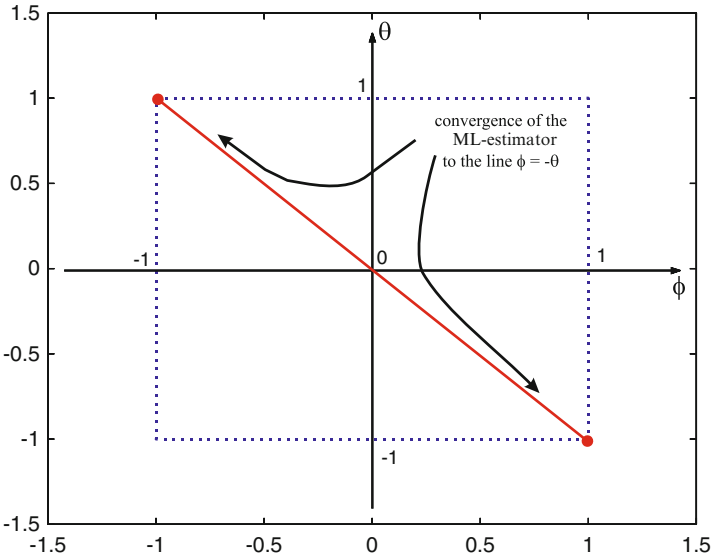

Fig. 5.1 Parameter space of a causal and invertible ARMA(1,1) process

following example. Suppose that $\{X_t\}$ is a white noise process, i.e. $X_t = Z_t \sim \mathrm{WN}(0, \sigma^2)$, but we fit an ARMA(1,1) model given by $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$. Then, the maximum likelihood estimator does not converge to $\phi = \theta = 0$, but only to the line-segment $\phi = -\theta$ with $|\phi| < 1$ and $|\theta| < 1$. For values of ϕ and θ on this line-segment we have $\psi(z) = \theta(z)/\phi(z) = 1$. The maximum likelihood estimator converges to the true values of ψ_j, i.e. to the values $\psi_0 = 1$ and $\psi_j = 0$ for $j > 0$. The situation is depicted in Fig. 5.1. There it is shown that the estimator has a tendency to converge to the points $(-1, 1)$ and $(1, -1)$, depending on the starting values. This indeterminacy of the estimator manifest itself as a numerical problem in the optimization of the likelihood function. Thus models with similar roots for the AR and MA polynomials which are close in absolute value to the unit circle are probably overparametrized. The problem can be overcome by reducing the orders of the AR and MA polynomial by one.

This problem does not appear in a purely autoregressive models. As explained in section "Example: AR(1) Process", the estimator for the redundant coefficients converges to zero with asymptotic distribution $\mathrm{N}(0, 1/T)$ (see the result in Eq. (5.1)). This is one reason why purely autoregressive models are often preferred. In addition the estimator is easily implemented and every stationary stochastic process can be arbitrarily well approximated by an AR process. This approximation may, however, necessitate high order models when the true process encompasses a MA component.

In the case of underfitting the maximum likelihood estimator converges to those values which are closest to the true parameters given the restricted parameter space. The estimates are, however, inconsistent due to the "omitted variable bias".

For these reasons the identification of the orders is an important step. One method which goes back to Box and Jenkins (1976) consists in the analysis of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) (see Sect. 3.5). Although this method requires some experience, especially when the process is not a purely AR or MA process, the analysis of the ACF und PACF remains an important first step in every practical investigation of a time series.

An alternative procedure relies on the automatic order selection. The objective is to minimize a so-called *information criterion* over different values of p and q. These criteria are based on the following consideration. Given a fixed number of observations, the successive increase of the orders p and q increases the fit of the model so that variance of the residuals $\hat{\sigma}^2_{p,q}$ steadily decreases. In order to compensate for this tendency to overfitting a penalty is introduced. This penalty term depends on the number of free parameters and on the number of observations at hand.[6] The most important information criteria have the following additive form:

$$\ln \hat{\sigma}^2_{p,q} \; + \; (\#\text{ free parameters}) \, \frac{C(T)}{T} \quad = \quad \ln \hat{\sigma}^2_{p,q} \; + \; (p \; + \; q) \frac{C(T)}{T} \quad \longrightarrow \quad \min_{p,q},$$

where $\ln \hat{\sigma}^2_{p,q}$ measures the goodness of fit of the ARMA(p,q) model and $(p+q)\frac{C(T)}{T}$ denotes the penalty term. Thereby $C(T)$ represents a nondecreasing function of T which governs the trade-off between goodness of fit and complexity (dimension) of the model. Thus, the information criteria chooses higher order models for larger sample sizes T. If the model includes a constant term or other exogenous variables, the criterion must be adjusted accordingly. However, this will introduce, for a given sample size, just a constant term in the objective function and will therefore not influence the choice of p and q.

The most common criteria are the Akaike information criterion (AIC), the Schwarz or Bayesian information criterion (BIC), and the Hannan-Quinn information criterion (HQ criterion):

$$\begin{aligned}
\text{AIC}(p,q) &= \ln \hat{\sigma}^2_{p,q} + (p+q)\frac{2}{T} \\
\text{BIC}(p,q) &= \ln \hat{\sigma}^2_{p,q} + (p+q)\frac{\ln T}{T} \\
\text{HQC}(p,q) &= \ln \hat{\sigma}^2_{p,q} + (p+q)\frac{2\ln(\ln T)}{T}
\end{aligned}$$

Because AIC $<$ HQC $<$ BIC for a given sample size $T \geq 16$, Akaike's criterion delivers the largest models, i.e. the highest order $p+q$; the Bayesian criterion is more restrictive and delivers therefore the smallest models, i.e. the lowest $p+q$. Although Akaike's criterion is not consistent with respect to p and q and has a

[6] See Brockwell and Davis (1991) for details and a deeper appreciation.

tendency to deliver overfitted models, it is still widely used in practice. This feature is sometimes desired as overfitting is seen as less damaging than underfitting.[7] Only the BIC and HQC lead to consistent estimates of the orders p and q.

5.5 Modeling a Stochastic Process

The identification of a satisfactory ARMA model typically involves in practice several steps.

Step 1: Transformations to Achieve Stationary Time Series

Economic time series are often of a non-stationary nature. It is therefore necessary to transform the time series in a first step to achieve stationarity. Time series which exhibit a pronounced trend (GDP, stock market indices, etc.) should not be modeled in levels, but in differences. If the variable under consideration is already in logs, as is often case, then taking first differences effectively amounts to working with growth rates. Sometimes first differences are not enough and further differences have to be taken. Price indices or monetary aggregates are typical examples where first differences may not be sufficient to achieve stationarity. Thus instead of X_t one works with the series $Y_t = (1 - \mathrm{L})^d X_t$ with $d = 1, 2, \ldots$. A non-stationary process $\{X_t\}$ which needs to be differentiated d-times to arrive at a stationary time series is called *integrated of order d*, $X_t \sim I(d)$.[8] If $Y_t = (1 - \mathrm{L})^d X_t$ is generated by an ARMA(p,q) process, $\{X_t\}$ is said to be an ARIMA(p,d,q) process.

An alternative method to eliminate the trend is to regress the time series against a polynomial in t of degree s, i.e. against $(1, t, \ldots, t^s)$, and to proceed with the residuals. These residuals can then be modeled as an ARMA(p,q) process. Chapter 7 discusses in detail which of the two detrending methods is to be preferred under which circumstances.

Often the data are subject to seasonal fluctuations. As with the trend there are several alternative available. The first possibility is to pass the time series through a seasonal filter and work with the seasonally adjusted data. The construction of seasonal filters is discussed in Chap. 6. A second alternative is to include seasonal dummies in the ARMA model. A third alternative is to take seasonal differences. In the case of quarterly observations, this amounts to work with $Y_t = (1 - \mathrm{L}^4)X_t$. As $1 - \mathrm{L}^4 = (1 - \mathrm{L})(1 + \mathrm{L} + \mathrm{L}^2 + \mathrm{L}^3)$, this transformation involves a first difference and will therefore also account for the trend.

[7] See, for example, the section on the unit root tests 7.3.

[8] An exact definition will be provided in Chap. 7. In this chapter we will analyze the consequences of non-stationarity and discuss tests for specific forms of non-stationarity.

Step 2: Finding the Orders p and q

Having achieved stationarity, one has to find the appropriate orders p and q of the ARMA model. Thereby one can rely either on the analysis of the ACF and the PACF, or on the information criteria outlined in the previous Sect. 5.4.

Step 3: Checking the Plausibility

After having identified a particular model or a set of models, one has to inspect its adequacy. There are several dimensions along which the model(s) can be checked.

(i) Are the residuals white noise? This can be checked by investigating at the ACF of the residuals and by applying the Ljung-Box test (4.4). If they are not this means that the model failed to capture all the dynamics inherent in the data.
(ii) Are the parameters plausible?
(iii) Are the parameters constant over time? Are there structural breaks? This can be done by looking at the residuals or by comparing parameter estimates across subsamples. More systematic approaches are discussed in Perron (2006). These involve the revolving estimation of parameters by allowing the break point to vary over the sample. Thereby different type of structural breaks can be distinguished. A more in depth analysis of structural breaks is presented in Sect. 18.1.
(iv) Does the model deliver sensible forecasts? It is particularly useful to investigate the out-of-sample forecasting performance. If one has several candidate models, one can perform a horse-race among them.

In case the model turns out to be unsatisfactory, one has to go back to steps 1 and 2.

5.6 An example: Modeling Real GDP in the Case of Switzerland

This section illustrates the concepts and ideas just presented by working out a specific example. We take the seasonally unadjusted Swiss real GDP as an example. The data are plotted in Fig. 1.3. To take the seasonality into account we transform the logged time series by taking first seasonal differences, i.e. $X_t = (1-\mathrm{L}^4)\ln \mathrm{GDP}_t$. Thus, the variable corresponds to the growth rate with respect to quarter of the previous year. The data are plotted in Fig. 5.2. A cursory inspection of the plot reveals that this transformation eliminated the trend as well as the seasonality.

First we analyze the ACF and the PACF. They are plotted together with corresponding confidence intervals in Fig. 5.3. The slowly monotonically declining ACF suggests an AR process. As only the first two orders of the PACF are significantly different from zero, it seems that an AR(2) model is appropriate. The least-squares estimate of this model are:

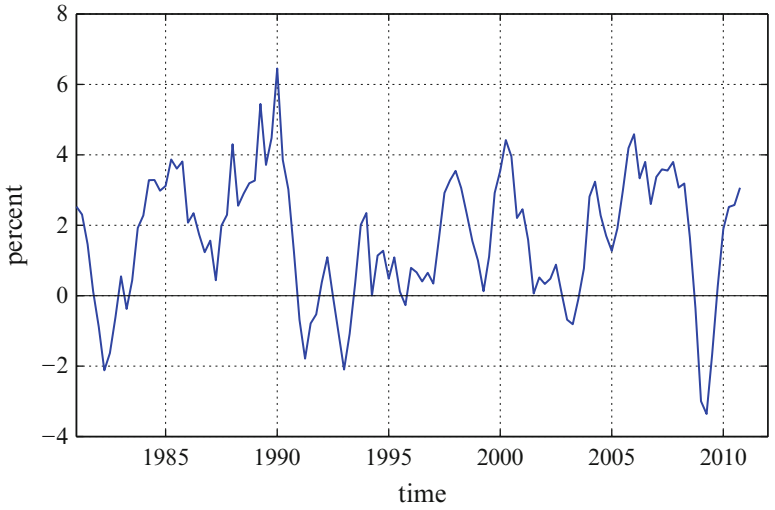

Fig. 5.2 Real GDP growth rates of Switzerland

$$X_t - \underset{(0,103)}{1.134}\ X_{t-1} + \underset{(0.104)}{0.310}\ X_{t-2} = 0.218 + Z_t \qquad \text{with } \hat{\sigma}^2 = 0.728$$

The numbers in parenthesis are the estimated standard errors of the corresponding parameter above. The roots of the AR-polynomial are 1.484 and 2.174. They are clearly outside the unit circle so that there exists a stationary and causal representation.

Next, we investigate the information criteria AIC and BIC to identify the orders of the ARMA(p,q) model. We examine all models with $0 \le p, q \le 4$. The AIC and the BIC values, are reported in Tables 5.1 and 5.2. Both criteria reach a minimum at $(p, q) = (1, 3)$ (bold numbers) so that both criteria prefer an ARMA(1,3) model. The parameters of this models are as follows:

$$X_t - \underset{(0.134)}{0.527}\ X_{t-1} = 0.6354 + Z_t + \underset{(0.1395)}{0.5106}\ Z_{t-1}$$
$$+ \underset{(0.1233)}{0.5611}\ Z_{t-2} + \underset{(0.1238)}{0.4635}\ Z_{t-3} \qquad \text{with } \hat{\sigma}^2 = 0.648.$$

The estimated standard errors of the estimated parameters are again reported in parenthesis below. The AR(2) model is not considerably worse than the ARMA(1,3) model, according to the BIC criterion it is even the second best model.

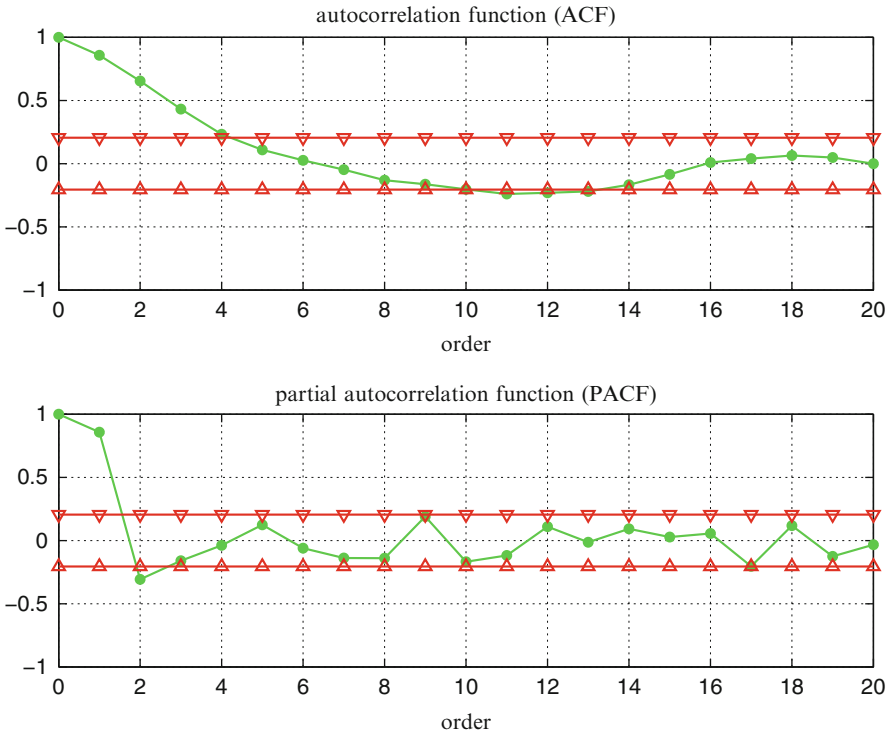

Fig. 5.3 Autocorrelation (ACF) and partial autocorrelation (PACF) function (PACF) of real GDP growth rates of Switzerland with 95 % confidence interval

Table 5.1 Values of Akaike's information criterium (AIC) for alternative ARMA(p,q) models

	q				
p	0	1	2	3	4
0		0.3021	0.0188	−0.2788	−0.3067
1	−0.2174	−0.2425	−0.2433	**−0.3446**	−0.2991
2	−0.2721	−0.2639	−0.2613	−0.3144	−0.2832
3	−0.2616	−0.2276	−0.2780	−0.2663	−0.2469
4	−0.2186	−0.1990	−0.2291	−0.2574	−0.2099

Minimum in bold

The inverted roots of the AR- and the MA-polynomial are plotted together with their corresponding 95 % confidence regions in Fig. 5.4.[9] As the confidence regions are all inside the unit circle, also the ARMA(1,3) has a stationary and causal representation. Moreover, the estimated process is also invertible. In addition, the roots of the AR- and the MA-polynomial are distinct.

[9]The confidence regions are determined by the delta-method (see Appendix E).

Table 5.2 Values of Bayes' information criterium (BIC) for alternative ARMA(p,q) models

p \ q	0	1	2	3	4
0		0.3297	0.0740	−0.1961	−0.1963
1	−0.1896	−0.1869	−0.1600	**−0.2335**	−0.1603
2	−0.2162	−0.1801	−0.1495	−0.1746	−0.1154
3	−0.1772	−0.1150	−0.1373	−0.0974	−0.0499
4	−0.1052	−0.0573	−0.0591	−0.0590	0.0169

Minimum in bold

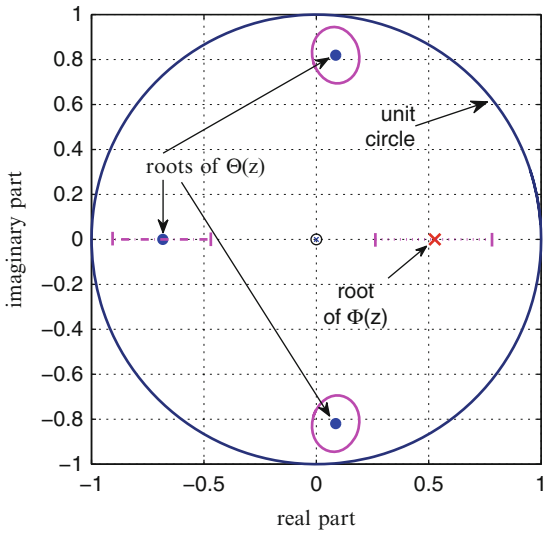

Fig. 5.4 Inverted roots of the AR- and the MA-polynomial of the ARMA(1,3) model together with the corresponding 95 % confidence regions

The autocorrelation functions of the AR(2) and the ARMA(1,3) model are plotted in Fig. 5.5. They show no sign of significant autocorrelations so that both residual series are practically white noise. We can examine this hypothesis formally by the Ljung-Box test (see Sect. 4.2 Eq. (4.4)). Taking N = 20 the values of the test statistics are $Q'_{\mathrm{AR(2)}} = 33.80$ and $Q'_{\mathrm{ARMA(1,3)}} = 21.70$, respectively. The 5 % critical value according to the χ^2_{20} distribution is 31.41. Thus the null hypothesis $\rho(1) = \ldots = \rho(20) = 0$ is rejected for the AR(2) model, but not for the ARMA(1,3) model. This implies that the AR(2) model does not capture the full dynamics of the data.

Although the AR(2) and the ARMA(1,3) model seem to be quite different at first glance, they deliver similar impulse response functions as can be gathered from Fig. 5.6. In both models, the impact of the initial shock is first built up to values higher than 1.1 in quarters one and two, respectively. Then the effect monotonically declines to zero. After 10 to 12 quarters the effect of the shock has practically dissipated.

As a final exercise, we use both models to forecast real GDP growth over the next nine quarters, i.e. for the period fourth quarter 2003 to fourth quarter 2005. As can

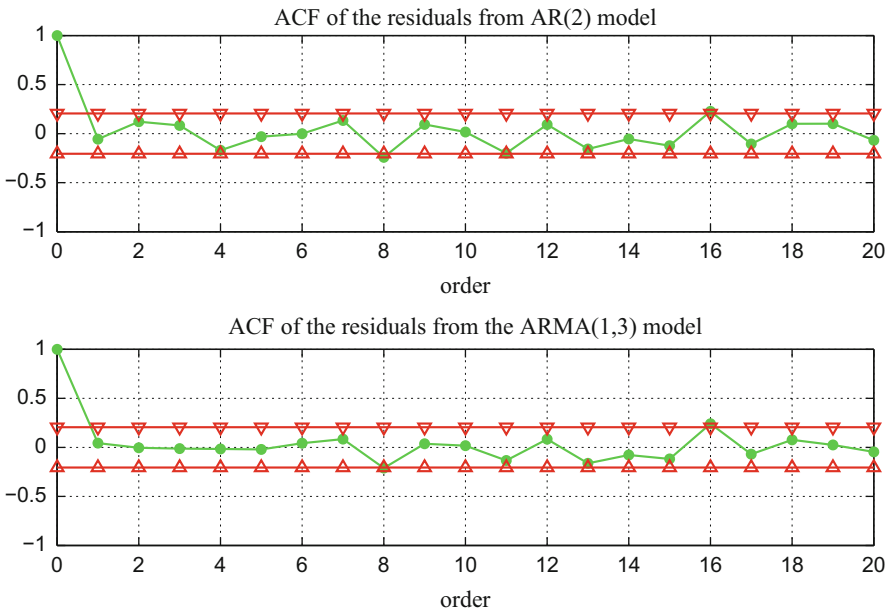

Fig. 5.5 Autocorrelation function (ACF) of the residuals from the AR(2) and the ARMA(1,3) model

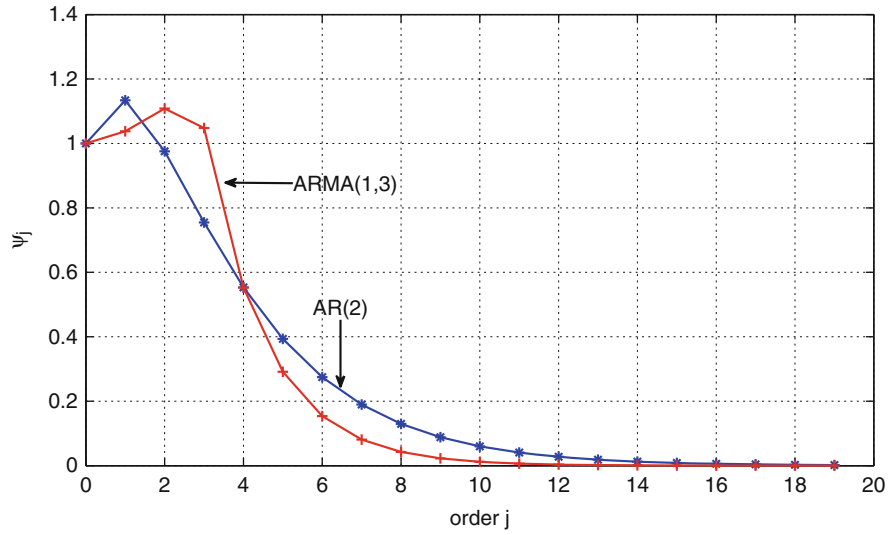

Fig. 5.6 Impulse responses of the AR(2) and the ARMA(1,3) model

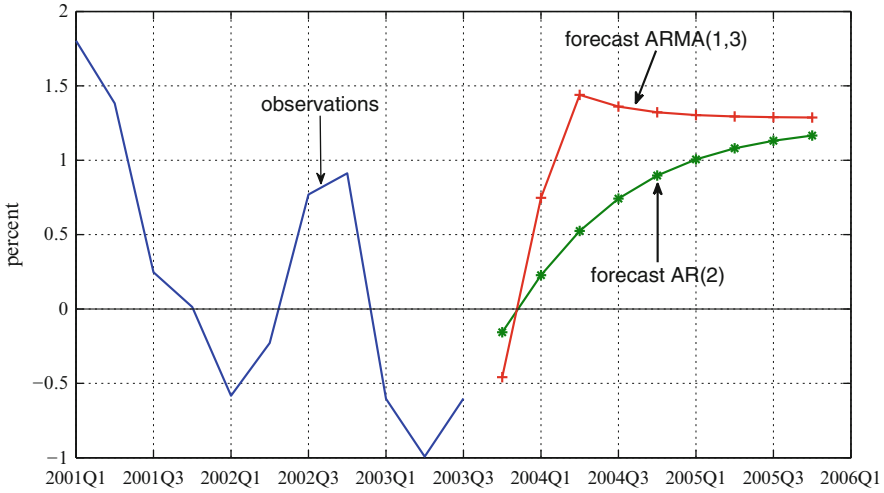

Fig. 5.7 Forecasts of real GDP growth rates for Switzerland

be seen from Fig. 5.7, both models predict that the Swiss economy should move out of recession in the coming quarters. However, the ARMA(1,3) model indicates that the recovery is taking place more quickly and the growth overshooting its long-run mean of 1.3 % in about a year. The forecast of the AR(2) predicts a more steady approach to the long-run mean.