



## Forecasting with Time Series Models

Having selected a model and fitted its parameters to a given times series, the model can then be used to estimate new data of the time series. If such data are estimated for a time period following the final data value  $X_T$  of the given time series, we speak of a *prediction* or *forecast*. The estimation of data lying between given data points is called *interpolation*. The question now arises as to how a model such as those given in Eqs. 32.6 or 32.13 could be used to obtain an “optimal” estimate. To answer this question the *forecasting error*

$$X_{T+k} - \widehat{X}_{T+k}, \quad k \in \mathbb{N}$$

between the estimated values  $\widehat{X}_{T+k}$  and the actual observed time series values  $X_{T+k}$  can be used if the last value used in the calibration of the model was  $X_T$ . The best forecast is that which minimizes the *mean square error* (MSE for short). The MSE is defined as the expectation of the squared forecasting error

$$\text{MQF} := E[(X_{T+k} - \widehat{X}_{T+k})^2]. \quad (33.1)$$

This expression is the mathematical formulation of the intuitive concept of the “distance” between the estimated and the actual values which is to be minimized on average (more cannot be expected when dealing with random variables). Minimizing this mean square error yields the result that the best forecasting estimate (called the *optimal forecast*) is given by the conditional expectation

$$\widehat{X}_{T+k} = E[X_{T+k} | X_T, \dots, X_2, X_1]. \quad (33.2)$$

This is the expectation of  $X_{T+k}$ , conditional on all available information about the time series up to and including  $T$ .

In practice, however, the concrete computation of this conditional expectation is generally very difficult since the joint distribution of the random variables must be known. Therefore, we often limit our consideration to the *linear forecast*

$$\widehat{X}_{T+k} = u_1 X_1 + u_2 X_2 + \cdots + u_T X_T \quad (33.3)$$

with appropriate coefficients  $u_i$ . This linear forecast is, in contrast to the optimal forecast, often more easily interpreted. For the special case that the  $\{X_i\}$  are normally distributed, the linear forecast and the optimal forecast agree. The *best* linear forecast can be characterized by the fact that the forecasting error  $X_{T+k} - \widehat{X}_{T+k}$  and the  $X_1, X_2, \dots, X_T$  are uncorrelated. The intuitive interpretation is that the  $X_1, X_2, \dots, X_T$  cannot provide any additional information for the forecast and the error is thus purely random.

### 33.1 Forecasting with Autoregressive Models

This forecasting procedure will now be applied to an AR( $p$ ) process, Eq. 32.6. The optimal one-step forecast is, according to Eq. 33.2, given directly by the conditional expectation in Eq. 32.7

$$\widehat{X}_{T+1}^{\text{optimal}} = E[X_{T+1}|X_T, \dots, X_1] = \sum_{i=1}^p \phi_i X_{T+1-i} . \quad (33.4)$$

This has the form indicated in Eq. 33.3. The optimal one-step forecast is thus the *best linear* one-step forecast. Equation 32.6 shows that the forecasting error  $X_{T+1} - \widehat{X}_{T+1}$  is precisely  $\varepsilon_{T+1}$  and independent of  $X_1, X_2, \dots, X_T$ . The MSE for the one-step forecast is given by Eq. 33.1 with  $k = 1$ . Thus,

$$\text{MSE} = E[(X_{T+1} - \widehat{X}_{T+1})^2] = E[\varepsilon_{T+1}^2] = \sigma^2 .$$

The optimal two-step forecast is the conditional expectation of  $X_{T+2}$  on the basis of knowledge of  $X_T, \dots, X_1$ , as can be seen from Eq. 33.2:

$$\widehat{X}_{T+2}^{\text{optimal}} = E[X_{T+2}|X_T, \dots, X_1] .$$

This is *not* the equivalent to the conditional expectation of  $X_{T+2}$  on the basis of knowledge of  $X_{T+1}, \dots, X_1$ . Hence, Eq. 32.7 cannot be applied directly. An additional difficulty arises due to the fact that  $X_{T+1}$  is unknown. The *optimal* two-step forecast cannot be calculated. We therefore proceed by computing the *linear* two-step forecast. The best linear two-step forecast is obtained by actually calculating the conditional expectation of  $X_{T+2}$  as if all the  $X_{T+1}, \dots, X_1$  were known and replacing the (unknown) value  $X_{T+1}$  by its best estimate  $\widehat{X}_{T+1}$  (which was calculated in the previous step):

$$\widehat{X}_{T+2}^{\text{linear}} = E[X_{T+2} | \widehat{X}_{T+1}, X_T, \dots, X_1].$$

Now Eq. 32.7 can be applied to this conditional expectation and utilizing Eq. 33.4 we obtain

$$\begin{aligned} \widehat{X}_{T+2}^{\text{linear}} &= \phi_1 \widehat{X}_{T+1} + \sum_{j=2}^p \phi_j X_{T+2-j} = \phi_1 \sum_{i=1}^p \phi_i X_{T+1-i} + \sum_{j=2}^p \phi_j X_{T+1-(j-1)} \\ &= \phi_1 \sum_{i=1}^p \phi_i X_{T+1-i} + \sum_{i=1}^{p-1} \phi_{i+1} X_{T+1-i} \\ &= \phi_1 \phi_p X_{T+1-p} + \sum_{i=1}^{p-1} [\phi_1 \phi_i + \phi_{i+1}] X_{T+1-i}. \end{aligned}$$

The linear two-step forecast then has the form indicated in Eq. 33.3. The forecasting error is found to be

$$\begin{aligned} X_{T+2} - \widehat{X}_{T+2} &= \underbrace{\sum_{i=1}^p \phi_i X_{T+2-i}}_{X_{T+2}, \text{ see Eqn. 32.6}} + \varepsilon_{T+2} - \phi_1 \widehat{X}_{T+1} - \sum_{j=2}^p \phi_j X_{T+2-j} \\ &= \varepsilon_{T+2} + \phi_1 (X_{T+1} - \widehat{X}_{T+1}) \\ &= \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}. \end{aligned}$$

Thus, the forecasting error is a sum of two normally distributed random variables and therefore itself normally distributed. The MSE can now be

computed as follows:

$$\begin{aligned} \text{MSE} &= E[(X_{T+2} - \widehat{X}_{T+2})^2] = E[(\varepsilon_{T+2} + \phi_1 \varepsilon_{T+1})^2] \\ &= E[\varepsilon_{T+2}^2] + \phi_1^2 E[\varepsilon_{T+1}^2] + 2\phi_1 E[\varepsilon_{T+2} \varepsilon_{T+1}] \\ &= \text{var}[\varepsilon_{T+2}^2] + \phi_1^2 \text{var}[\varepsilon_{T+1}^2] + 2\phi_1 0 \\ &= \sigma^2(1 + \phi_1^2) . \end{aligned}$$

This implies that the forecasting error  $X_{T+2} - \widehat{X}_{T+2}$  of the two-step forecast is normally distributed with variance  $\sigma^2(1 + \phi_1^2)$ , i.e.,  $N(0, \sigma^2(1 + \phi_1^2))$ .

Proceeding analogously, the best linear  $h$ -step forecast is obtained by taking the conditional expectation of  $X_{T+h}$  as if all  $X_t$  were known up to  $X_{T+h-1}$  and then replacing the yet unknown values of  $X_t$  for  $T < t < h$  with their best estimators calculated inductively in previous steps as described above:

$$\widehat{X}_{T+h}^{\text{linear}} = E[X_{T+h} | \widehat{X}_{T+h-1}, \widehat{X}_{T+h-2}, \dots, \widehat{X}_{T+1}, X_T, \dots, X_1] .$$

Equation 32.7 is then applied to these conditional expectations resulting in

$$\widehat{X}_{T+h} = \sum_{i=1}^{\min(h-1, p)} \phi_i \widehat{X}_{T+h-i} + \sum_{j=h}^p \phi_j X_{T+h-j} .$$

The forecasting error of the  $h$ -step forecast is

$$\begin{aligned} X_{T+h} - \widehat{X}_{T+h} &= \underbrace{\sum_{i=1}^p \phi_i X_{T+h-i} + \varepsilon_{T+h}}_{X_{T+h}, \text{ see Eqn. 32.6}} - \sum_{i=1}^{\min(h-1, p)} \phi_i \widehat{X}_{T+h-i} - \sum_{j=h}^p \phi_j X_{T+h-j} \\ &= \varepsilon_{T+h} + \sum_{i=1}^{\min(h-1, p)} \phi_i X_{T+h-i} + \sum_{i=h}^p \phi_i X_{T+h-i} \\ &\quad - \sum_{i=1}^{\min(h-1, p)} \phi_i \widehat{X}_{T+h-i} - \sum_{j=h}^p \phi_j X_{T+h-j} \\ &= \varepsilon_{T+h} + \sum_{i=1}^{\min(h-1, p)} \phi_i (X_{T+h-i} - \widehat{X}_{T+h-i}) . \end{aligned}$$

This is a recursion expressing the  $h$ -step forecasting error in terms of the forecasting errors for fewer than  $h$  steps. From this it can be shown that the  $h$ -step forecasting error is distributed as  $N(0, \sigma^2(1 + \phi_1^2 + \dots + \phi_{h-1}^2))$ .

The unknown coefficients  $\phi_1, \phi_2, \dots, \phi_p$  are estimated from the time series as shown in Sect. 32.2.1. The  $\phi_i$  in the forecast equation are simply replaced with  $\widehat{\phi}_i$ .

## 33.2 Volatility Forecasts with GARCH( $p, q$ ) Processes

GARCH models of the form indicated in Eq. 32.13 are not suitable for the prediction of the actual values  $X_i$  of a time series since the random variable in Eq. 32.13 appears as a *product* (rather than a sum as in Eq. 32.6). In consequence, the conditional expectations of the  $X_i$  are identically zero. However, GARCH models are well adapted for forecasting the (conditional) *variance* of time series values. According to Eq. 33.2, the conditional expectation is in general the optimal forecast. We are therefore looking for the *conditional expectation of the conditional variance*.

### 33.2.1 Forecast Over Several Time Steps

#### The One-Step Forecast

Equation 32.16 shows that the conditional variance of  $X_T$  is equal to  $H_T$  if all  $X_t$  for  $t \leq T - 1$  are known. Its conditional expectation is then the conditional expectation of  $H_T$ . Based on Eq. 32.17, the conditional expectation of  $H_T$  is simply  $H_T$  itself if the  $X$  values are known up to the time  $T - 1$ . Hence, the *optimal* one-step forecast for the conditional variance is<sup>1</sup>

$$\begin{aligned}\widehat{\text{var}}_{T+1}^{\text{optimal}} &= E[\text{var}_{T+1} | X_T, \dots, X_1] \\ &= E[H_{T+1} | X_T, \dots, X_1]\end{aligned}$$

---

<sup>1</sup>In order to keep the notation as simple as possible, we will adopt the convention of denoting the conditional variance by  $\text{var}_{T+h} := \text{var}[X_{T+h} | X_T, \dots, X_1]$ , likewise for its estimators  $\widehat{\text{var}}_{T+h} := \widehat{\text{var}}[X_{T+h} | X_T, \dots, X_1]$  for any  $h > 0$ . When these abbreviations for the conditional variances are used, it is always to be understood that the values  $X_T, \dots, X_1$  are known.

$$\begin{aligned}
&= H_{T+1} \\
&= \alpha_0 + \sum_{j=1}^q \alpha_j X_{T+1-j}^2 + \sum_{i=1}^p \beta_i H_{T+1-i} . \tag{33.5}
\end{aligned}$$

To clarify the argument used in the derivation of this result, each of the equalities in Eq. 33.5 will receive somewhat more scrutiny. The first equation is obtained from the general forecast equation, Eq. 33.2. The second follows from Eq. 32.16. The third holds as a result of Eq. 32.17 while the fourth equation is derived from Eq. 32.13 used in the construction of the GARCH process.

### The Two-Step Forecast

The two-step forecast is somewhat more complicated. The optimal two-step forecast is, according to Eq. 33.2, the conditional expectation of  $\text{var}_{T+2}$  under the condition that  $X_T, \dots, X_1$  are known:

$$\widehat{\text{var}}_{T+2}^{\text{optimal}} = E[\text{var}_{T+2} | X_T, \dots, X_1] .$$

Again, this is *not* equal to the conditional expectation of  $\text{var}_{T+2}$  under the condition that  $X_{T+1}, \dots, X_1$  are known. Equation 32.16 cannot be applied directly. Indeed, the *optimal* two-step forecast cannot be computed. We calculate instead, analogously to the linear forecast of the AR( $p$ ) process illustrated in Sect. 33.1, the best possible two-step forecast by replacing the expectation of  $\text{var}_{T+2}$  conditional upon  $X_T, \dots, X_1$  with the conditional expectation of  $\text{var}_{T+2}$  as if the  $X_{T+1}, \dots, X_1$  were all known:

$$\widehat{\text{var}}_{T+2} = E[\text{var}_{T+2} | X_{T+1}, \dots, X_1] .$$

Now Eq. 32.16 can be applied to obtain

$$\widehat{\text{var}}_{T+2} = E[\text{var}_{T+2} | X_{T+1}, \dots, X_1] = H_{T+2} .$$

Remember however, that  $X_{T+1}$  is *not* known and therefore  $H_{T+2}$  appearing here is not known at time  $T$ . The best we can do is to replace  $H_{T+2}$  by its optimal estimator which, according to Eq. 33.2, is given by its conditional

expectation

$$\widehat{\text{var}}_{T+2} = E[H_{T+2}|X_T, \dots, X_1]. \quad (33.6)$$

Inside *this* expectation, we now replace  $H_{T+2}$  in accordance with the construction in 32.13:

$$\begin{aligned} \widehat{\text{var}}_{T+2} &= E[H_{T+2}|X_T, \dots, X_1] \\ &= \alpha_0 + \sum_{j=1}^q \alpha_j E[X_{T+2-j}^2|X_T, \dots, X_1] + \sum_{i=1}^p \beta_i E[H_{T+2-i}|X_T, \dots, X_1] \\ &= \alpha_0 + \alpha_1 E[X_{T+1}^2|X_T, \dots, X_1] + \sum_{j=2}^q \alpha_j X_{T+2-j}^2 + \sum_{i=1}^p \beta_i H_{T+2-i}. \end{aligned} \quad (33.7)$$

In the last step, we have exploited the fact that all  $X_t$  are known for all times  $t \leq T$  and, according to Eq. 32.17, all of the  $H_t$  for the times  $t \leq T + 1$ . The expectation of the known quantities can be replaced by the quantities themselves. Only one unknown quantity remains, namely  $X_{T+1}^2$ . Because of Eq. 32.16, the conditional expectation of all of the  $X_{T+h}$  for all  $h > 0$  is always equal to zero. The expectation of  $X_{T+h}^2$  can therefore be replaced by the variance of  $X_{T+h}$ :

$$\begin{aligned} E[X_{T+h}^2|X_T, \dots, X_1] &= E[X_{T+h}^2|X_T, \dots, X_1] - \underbrace{(E[X_{T+h}|X_T, \dots, X_1])^2}_0 \\ &= \text{var}[X_{T+h}|X_T, \dots, X_1] \quad \text{for every } h > 0. \end{aligned} \quad (33.8)$$

For  $h = 1$  this implies

$$E[X_{T+1}^2|X_T, \dots, X_1] = \text{var}[X_{T+1}|X_T, \dots, X_1] = H_{T+1},$$

where Eq. 32.16 has again been used in the last step. The two-step forecast then becomes

$$\widehat{\text{var}}_{T+2} = \alpha_0 + \alpha_1 H_{T+1} + \sum_{j=2}^q \alpha_j X_{T+2-j}^2 + \sum_{i=1}^p \beta_i H_{T+2-i}. \quad (33.9)$$

### The Three-Step Forecast

For the two-step forecast, only the value of  $X_{T+1}$  in Eq. 33.7 was unknown, the necessary  $H$  values were known up to time  $T + 1$ . This is no longer the case in the three-step forecast. In this case, some of the  $H$  values are also unknown. Because of this additional difficulty, it is advisable to demonstrate the computation of a three-step forecast before generalizing to arbitrarily many steps.

The three-step forecast now proceeds analogous to the two-step forecast: the optimal forecast is, as indicated in Eq. 33.2, the conditional expectation of  $X_{T+3}$  under the condition that  $X_T, \dots, X_1$  are known.

$$\widehat{\text{var}}_{T+3}^{\text{optimal}} = E[\text{var}_{T+3} | X_T, \dots, X_1] .$$

Again, Eq. 32.16 cannot be directly applied since the  $X$  are only known up to  $X_T$  and not up to  $X_{T+2}$ . The best possible three-step forecast is thus, analogous to Eq. 33.6

$$\widehat{\text{var}}_{T+3} = E[H_{T+3} | X_T, \dots, X_1] . \tag{33.10}$$

In this expectation we now replace  $H_{T+3}$  with its expression constructed in Eq. 32.13 to obtain

$$\begin{aligned} \widehat{\text{var}}_{T+3} &= E[H_{T+3} | X_T, \dots, X_1] \\ &= \alpha_0 + \sum_{i=1}^p \beta_i E[H_{T+3-i} | X_T, \dots, X_1] + \sum_{j=1}^q \alpha_j E[X_{T+3-j}^2 | X_T, \dots, X_1] \\ &= \alpha_0 + \sum_{i=2}^p \beta_i H_{T+3-i} + \beta_1 E[H_{T+2} | X_T, \dots, X_1] \\ &\quad + \alpha_1 E[X_{T+2}^2 | X_T, \dots, X_1] + \alpha_2 E[X_{T+1}^2 | X_T, \dots, X_1] + \sum_{j=3}^q \alpha_j X_{T+3-j}^2 . \end{aligned} \tag{33.11}$$

In the last step, the expectations of the known values were again replaced by the values themselves (all  $X_t$  for times  $t \leq T$  and all  $H_t$  for times  $t \leq T + 1$ ). Only three unknown values remain, namely  $X_{T+1}^2, X_{T+2}^2$  and  $H_{T+2}$ . For the



conditional expectation of  $X_{T+1}^2$  and  $X_{T+2}^2$  we can use Eq. 33.8 to write

$$\begin{aligned} E[X_{T+1}^2 | X_T, \dots, X_1] &= \text{var}[X_{T+1} | X_T, \dots, X_1] = H_{T+1} \\ E[X_{T+2}^2 | X_T, \dots, X_1] &= \text{var}[X_{T+2} | X_T, \dots, X_1] = \widehat{\text{var}}_{T+2} . \end{aligned}$$

Equation 32.16 has been used in the first of the above two equations. In the second equation, this is not possible since taking the conditional variance at time  $T+2$  under the condition that  $X_T, \dots, X_1$  are known is not the same as taking it conditional upon knowing the values of  $X_{T+1}, \dots, X_1$ . We have no other choice than to replace the unknown  $\text{var}[X_{T+2} | X_T, \dots, X_1]$  with the (previously calculated) estimator  $\widehat{\text{var}}_{T+2}$ .

For the expectation  $E[H_{T+2} | X_T, \dots, X_1]$  we make use of the fact that, according to Eq. 33.6, it is equal to the two-step forecast for the variance

$$E[H_{T+2} | X_T, \dots, X_1] = \widehat{\text{var}}_{T+2} .$$

Substituting all this into Eq. 33.11 finally yields

$$\widehat{\text{var}}_{T+3} = \alpha_0 + (\alpha_1 + \beta_1) \widehat{\text{var}}_{T+2} + \alpha_2 H_{T+1} + \sum_{j=3}^q \alpha_j X_{T+3-j}^2 + \sum_{i=2}^p \beta_i H_{T+3-i} . \quad (33.12)$$

### The Forecast for $h$ Steps

The generalization to the forecast for an arbitrary number of steps  $h$  is now quite simple. Analogous to Eqs. 33.6 and 33.10 the best possible estimate is

$$\widehat{\text{var}}_{T+h} = E[H_{T+h} | X_T, \dots, X_1] \quad \text{for every } h > 0 . \quad (33.13)$$

Within this expectation, we now replace  $H_{T+h}$  as in the construction Eq. 32.13 and obtain an equation analogous to Eq. 33.11

$$\begin{aligned} \widehat{\text{var}}_{T+h} &= E[H_{T+h} | X_T, \dots, X_1] \\ &= \alpha_0 + \sum_{j=1}^q \alpha_j E[X_{T+h-j}^2 | X_T, \dots, X_1] + \sum_{i=1}^p \beta_i E[H_{T+h-i} | X_T, \dots, X_1] \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=1}^{\min(h-1, q)} \alpha_j E[X_{T+h-j}^2 | X_T, \dots, X_1] + \sum_{j=h}^q \alpha_j X_{T+h-j}^2 + \alpha_0 \\
 &+ \sum_{i=1}^{\min(h-2, p)} \beta_i E[H_{T+h-i} | X_T, \dots, X_1] + \sum_{i=h-1}^p \beta_i H_{T+h-i} .
 \end{aligned}$$

In the last step, the expectation of the known values have again been replaced by the values themselves (all  $X_t$  for times  $t \leq T$  and all  $H_t$  for times  $t \leq T+1$ ). The remaining expectations of the  $H$ s are replaced according to Eq. 33.13 through the respective variance estimators. We again use Eq. 33.8 for the conditional expectation of  $X^2$  and write

$$E[X_{T+1}^2 | X_T, \dots, X_1] = \text{var}[X_{T+1} | X_T, \dots, X_1] = H_{T+1}$$

$$E[X_{T+k}^2 | X_T, \dots, X_1] = \text{var}[X_{T+k} | X_T, \dots, X_1] = \widehat{\text{var}}_{T+k} \quad \text{for } k > 1 .$$

Substituting all of these relations for the conditional expectations finally yields the general  $h$ -step forecast of the conditional volatility in the GARCH( $p, q$ ) model:

$$\begin{aligned}
 \widehat{\text{var}}_{T+h} &= \sum_{j=1}^{\min(h-2, q)} \alpha_j \widehat{\text{var}}_{T+h-j} + \sum_{j=h}^q \alpha_j X_{T+h-j}^2 + \alpha_{h-1} H_{T+1} + \alpha_0 \\
 &+ \sum_{i=1}^{\min(h-2, p)} \beta_i \widehat{\text{var}}_{T+h-i} + \sum_{i=h-1}^p \beta_i H_{T+h-i} .
 \end{aligned} \tag{33.14}$$

Together with the start value, Eq. 33.5, in the form of  $\widehat{\text{var}}_{T+1} = H_{T+1}$  the  $h$ -step forecast can be computed recursively for all  $h$ .

From Eqs. 33.13 and 33.2, the estimator for the variance is simultaneously the estimator for  $H$ , and thus

$$\widehat{H}_{T+h} = \widehat{\text{var}}_{T+h} \quad \text{for all } h > 0 .$$

### 33.2.2 Forecast for the Total Variance

In the financial world, the time series  $\{X_T\}$  is usually taken to represent a relative price change (yield), see for example Fig. 32.1 of the FTSE data. We

are therefore also interested in the variance of the return over an *entire* time period of length  $h$ , i.e., in the variance of the sum  $\sum_{j=1}^h X_{T+j}$ . In forecasts such as in Eq. 33.14, we are only dealing with a prediction of the conditional variance *after*  $h$  steps and not with a prediction of the variance *over* the entire term of  $h$  steps (from  $X_T$  to  $X_{T+h}$ ). In other words, Eq. 33.14 is forecasting the conditional variance of  $X_{T+h}$  alone, and not predicting the variance of the sum  $\sum_{j=1}^h X_{T+j}$ . The variance of the total return  $\sum_{j=1}^h X_{T+j}$  for independent (in particular uncorrelated) returns is simply the sum of the variances as can be seen in Eq. A.17. Since the  $X_t$  of the process in Eq. 32.13 are uncorrelated (because the  $\varepsilon_t$  are *iid*), the estimator for the total variance of the GARCH process over  $h$  steps is simply

$$\widehat{\text{var}}\left[\sum_{j=1}^h X_{T+j} \mid X_T, \dots, X_1\right] = \sum_{j=1}^h \widehat{\text{var}}[X_{T+j} \mid X_T, \dots, X_1] = \sum_{j=1}^h \widehat{\text{var}}_{T+j} . \tag{33.15}$$

Even in the case of weak autocorrelations between the returns in a given time series, this result holds in good approximation.

### 33.2.3 Volatility Term Structure

The variance of the total return over a term from  $T$  until  $T + h$  is a function of this term. The square root of the (annualized) variance of the total return as a function of the term is called the *volatility term structure*. This plays an important role in pricing options since for an option with a lifetime of  $h$ , the volatility associated with this term is the relevant parameter value. From the estimator for the variance of the total return over the pertinent term, we obtain the estimator of the volatility structure as

$$\sigma(T, T + h) = \sqrt{\frac{1}{h} \widehat{\text{var}} \left[ \sum_{i=1}^h X_{T+i} \mid X_T, \dots, X_1 \right]} = \sqrt{\frac{1}{h} \sum_{j=1}^h \widehat{\text{var}}_{T+j}} . \tag{33.16}$$

### 33.3 Volatility Forecasts with GARCH (1,1) Processes

For the GARCH(1, 1) process ( $q = 1, p = 1$ ), all of the above estimators can be computed explicitly and the recursion equation 33.14 can be carried out. The start value of the recursion is simply

$$\widehat{\text{var}}_{T+1} = H_{T+1} = \alpha_0 + \alpha_1 X_T^2 + \beta_1 H_T . \quad (33.17)$$

as can be seen from Eq. 33.5. The two-step forecast as given by Eq. 33.9 simplifies to

$$\widehat{\text{var}}_{T+2} = \alpha_0 + \kappa H_{T+1}$$

where we defined the abbreviation

$$\kappa := \alpha_1 + \beta_1 .$$

For  $h > 2$  and  $p = q = 1$ , the upper limits in the sums over the variance estimators in the general recursion equation 33.14 are simply

$$\min(h - 2, q) = \min(h - 2, p) = 1 \quad \text{for } h > 2 .$$

Neither of the other sums makes any contribution since the lower limit in these sums is greater than the upper limit. The term  $\alpha_{h-1} H_{T+1}$  likewise does not exist since  $q = 1$  implies that only  $\alpha_0$  and  $\alpha_1$  exist. However,  $h - 1$  is greater than 1 for  $h > 2$ . All things considered, the  $h$ -step forecast in Eq. 33.14 reduces to

$$\widehat{\text{var}}_{T+h} = \alpha_0 + \kappa \widehat{\text{var}}_{T+h-1}$$

where  $\kappa = \alpha_1 + \beta_1$ . This recursion relation has a closed form expression in the form of a geometric series:

$$\begin{aligned} \widehat{\text{var}}_{T+h} &= \alpha_0 + \kappa \widehat{\text{var}}_{T+h-1} \\ &= \alpha_0 + \kappa(\alpha_0 + \kappa \widehat{\text{var}}_{T+h-2}) = \alpha_0(1 + \kappa) + \kappa^2 \widehat{\text{var}}_{T+h-2} \\ &= \alpha_0(1 + \kappa) + \kappa^2(\alpha_0 + \kappa \widehat{\text{var}}_{T+h-3}) = (\alpha_0(1 + \kappa + \kappa^2) + \kappa^3 \widehat{\text{var}}_{T+h-3}) \\ &\dots \end{aligned}$$

$$\begin{aligned}
&= \alpha_0 \underbrace{\sum_{i=1}^{h-1} \kappa^{i-1}}_{\text{geometric series}} + \kappa^{h-1} \underbrace{\widehat{\text{var}}_{T+1}}_{H_{T+1}} \\
&= \alpha_0 \left( \frac{1 - \kappa^{h-1}}{1 - \kappa} \right) + \kappa^{h-1} H_{T+1} \\
&= \tilde{\alpha}_0 + \kappa^{h-1} (H_{T+1} - \tilde{\alpha}_0) \\
&= \tilde{\alpha}_0 + \kappa^{h-1} \left( \alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \tilde{\alpha}_0 \right), \quad h > 1, \quad (33.18)
\end{aligned}$$

where for  $\widehat{\text{var}}_{T+1}$  the start-value of the recursion  $H_{T+1}$  was used and the geometric series was calculated according to Eq. 15.10. Here

$$\tilde{\alpha}_0 := \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

again denotes the *unconditional* variance from Eq. 32.18. The GARCH(1,1) prediction for the conditional variance after  $h$  steps is therefore equal to the unconditional variance plus the difference between the one-step forecast and the unconditional variance dampened by the factor  $\kappa^{h-1}$ . The stationarity condition requiring that  $\alpha_1 + \beta_1 < 1$  implies that for  $h \rightarrow \infty$  (a long prediction period) the GARCH prediction converges towards the unconditional variance.

The variance of the total return  $\sum_{i=1}^h X_{T+i}$  over a term of length  $h$  as the sum of the conditional forecasts is obtained for the *textGARCH*(1, 1) process as indicated in Eq. 33.15:

$$\begin{aligned}
\widehat{\text{var}}\left[\sum_{i=1}^h X_{T+i} \mid X_T, \dots, X_1\right] &= \sum_{i=1}^h \widehat{\text{var}}_{T+i} \\
&= \sum_{i=1}^h \left[ \tilde{\alpha}_0 + \kappa^{i-1} \left( \alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \tilde{\alpha}_0 \right) \right] \\
&= h\tilde{\alpha}_0 + \left( \frac{1 - \kappa^h}{1 - \kappa} \right) \left( \alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \tilde{\alpha}_0 \right), \quad (33.19)
\end{aligned}$$

where Eq. 15.10 for the geometric series is used again in the last step. The volatility term structure resulting from the 33.16 GARCH(1, 1) process is thus

$$\begin{aligned}\sigma(T, T+h) &= \sqrt{\frac{1}{h} \widehat{\text{var}}\left[\sum_{i=1}^h X_{T+i} | X_T, \dots, X_1\right]} \\ &= \sqrt{\tilde{\alpha}_0 + \frac{1}{h} \left(\frac{1-\kappa^h}{1-\kappa}\right) (\alpha_0 + \beta_1 H_T + \alpha_1 X_T^2 - \tilde{\alpha}_0)}.\end{aligned}$$

which approaches the unconditional variance  $\tilde{\alpha}_0$  like  $1/\sqrt{h}$  for large  $h$ .

### 33.4 Volatility Forecasts with Moving Averages

In addition to the relatively modern GARCH models, older methods such as *moving averages* exist in the market, which, despite their obvious shortcomings, are still widely used, thanks to their simplicity. Before entering into a discussion of volatility forecasts via moving averages and comparing them with those of the GARCH models, we will first introduce the two most important varieties, the *simple moving average*, abbreviated here as *MA* and the *exponentially weighted moving average*, abbreviated as *EWMA*.

The (simple) moving average measures the conditional variance (of a time series with zero mean) simply as the sum of evenly weighted squared time series values over a time window of width  $b$ . The form for the MA corresponding to Eq. 32.16 is simply

$$\text{Var}[X_t | X_{t-1}, \dots, X_1] = \frac{1}{b} \sum_{k=1}^b X_{t-k}^2.$$

The well-known *phantom structures* arise from this equation because every swing in the  $X_t^2$  is felt fully for  $b$  periods and then suddenly disappears completely when the term causing the perturbation no longer contributes to the average. An improvement would be to consider *weighted* sums where time series values further in the past are weighted less than values closer to the present. This can be realized, for example, by the exponentially weighted

moving average EWMA. The conditional variance in the EWMA is

$$\begin{aligned}\text{Var}[X_t|X_{t-1}, \dots, X_1] &= \frac{1}{\sum_{j=1}^b \lambda^{j-1}} \sum_{k=1}^b \lambda^{k-1} X_{t-k}^2 \\ &= \frac{X_{t-1}^2 + \lambda^1 X_{t-2}^2 + \dots + \lambda^{b-1} X_{t-b}^2}{1 + \lambda^1 + \lambda^2 + \dots + \lambda^{b-1}}.\end{aligned}$$

For  $\lambda < 1$  the values lying further back contribute less. The values commonly assigned to the parameter  $\lambda$  lie between 0.8 and 0.98. Naturally, the simple MA can be interpreted as a special case of the EWMA with  $\lambda = 1$ . The conditional variance of the EWMA is very similar to that of a GARCH(1,1) process since the recursion for  $H_T$  in Eq. 32.16 can be performed explicitly for  $p = q = 1$  and the conditional variance of the GARCH(1,1) process becomes

$$\begin{aligned}\text{Var}[X_t|X_{t-1}, \dots, X_1] &= H_t \\ &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 H_{t-1} \\ &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 [\alpha_0 + \beta_1 H_{t-2} + \alpha_1 X_{t-2}^2] \\ &= \alpha_0(1 + \beta_1) + \alpha_1(X_{t-1}^2 + \beta_1 X_{t-2}^2) + \beta_1^2 [\alpha_0 + \beta_1 H_{t-3} + \alpha_1 X_{t-3}^2] \\ &= \alpha_0(1 + \beta_1 + \beta_1^2) + \alpha_1(X_{t-1}^2 + \beta_1 X_{t-2}^2 + \beta_1^2 X_{t-3}^2) + \beta_1^3 H_{t-3} \\ &\dots \\ &= \alpha_0 \sum_{k=1}^b \beta_1^{k-1} + \alpha_1 \sum_k \beta_1^{k-1} X_{T-k}^2 + \beta_1^b H_{T-b}.\end{aligned}$$

If we now choose the parameters  $\alpha_0 = 0$ ,  $\beta_1 = \lambda$  and  $\alpha_1 = \lambda(\sum_{j=1}^b \lambda^j)^{-1}$  then this conditional variance after  $b$  steps (apart from remainder term  $\beta_1^b H_{T-b}$  which contains the influence of factors lying still further in the past) is exactly the same expression as for the EWMA. The difference between the GARCH(1, 1) and EWMA models first appears clearly in variance forecasts over more than one time step.

The conditional variances presented above can be interpreted as a one-step forecast for the conditional variance. The forecast over  $h$  steps delivers nothing new for the moving averages since both the MA and EWMA are static and fail to take the time structure into consideration. They start with the basic assumption that prices are lognormally distributed with a constant volatility.

This implies that

$$\widehat{\text{var}}[X_{T+h}|X_T, \dots, X_1] = \widehat{\text{var}}[X_{T+1}|X_T, \dots, X_1] = \text{var}[X_{T+1}|X_T, \dots, X_1].$$

holds for the  $h$ -step forecast of the MA as well as for the EWMA. As mentioned after Eq. 33.14, the conditional variance *after*  $h$  steps is being forecasted and *not* the variance of the total return *over* a term of  $h$  steps. The prediction for the variance of the total return as the sum over the conditional one-step forecasts is for moving averages (MA and EWMA) simply

$$\widehat{\text{var}}\left[\sum_{i=1}^h X_{T+i}|X_T, \dots, X_1\right] = \sum_{i=1}^h \underbrace{\widehat{\text{var}}[X_{T+i}|X_T, \dots, X_1]}_{\widehat{\text{var}}[X_{T+1}|X_T, \dots, X_1]} = h \text{var}[X_{T+1}|X_T, \dots, X_1].$$

This is again the famous *square root law* for the growth of the standard deviation over time. The variance simply increases linearly over time and the standard deviation is therefore proportional to the square root of time. This leads to a static prediction of the volatility, and extrapolating, for example, daily to yearly volatilities in this way can easily result in an overestimation of the volatilities. The *volatility term structure* for the moving average is then, as expected, a constant:

$$\sigma(T, T+h) = \sqrt{\frac{1}{h} \widehat{\text{var}}\left[\sum_{i=1}^h X_{T+i}|X_T, \dots, X_1\right]} = \sqrt{\text{var}[X_{T+1}|X_T, \dots, X_1]}.$$

In the Excel workbook GARCH.XLSX, the one-step forecast of a GARCH(1, 1) process, an MA with  $b = 80$  and an EWMA with  $b = 80$  and  $\lambda = 0,95$  are presented. Furthermore, the ten-step forecast of the GARCH(1, 1) process is shown. Since the time series we are dealing with is a simulated GARCH process, the “true” volatility is known (it is the  $H_t$  from the simulated series) and direct comparison can be made with each of the various estimates. As is clearly illustrated in GARCH.XLSX, the one-step GARCH(1, 1) forecast ( $H_t$  with the parameters  $\widehat{\alpha}_0$ ,  $\widehat{\alpha}_1$  and  $\widehat{\beta}_1$  fitted by *simulated annealing*) produces estimates which are quite close to the true volatility. The computation of the GARCH volatility term structure is presented in GARCH.XLSX as well.