Time Series

John Boland

August 8, 2022

Contents

List of Tables

List of Figures

1 Principles of statistical forecasting

- The main item refers not only to forecasting but any type of statistical modelling.
- It is that one builds the model one part (usually the majority) of the data, and then tests the results on the remainder.
- These sets are termed the **Training** and **Testing** sets.
- In standard statistical modelling, often the sets are sampled from the data at random.
- In time series forecasting, this is not possible, as the model building set has to have consecutive data, as does the testing set.
- This is further complicated with data that includes seasonality, as at least the training set has to contain multiples of the fundamental period.
- For instance, for hourly solar radiation, the training set would comprise a number of years. We also have to cater for leap years in some way. The easiest way is to delete Feb 29.

2 Example - Lake Huron

- In keeping with the idea of training and testing sets, I took the first 75 years as the training set, reserving 22 years for testing.
- I fitted a line through those 75 data points.
- I then took the residuals of that fit, and developed an $AR(2)$ model for them.
- Let $L(t)$ be the level of the lake, and $f(t)$ be the linear fit.
- Then $r(t) = L(t) - f(t)$ will be the residuals.

The models are

$$
\begin{aligned}
f(t) &= 10.64 - 0.041t \\
r(t) &= 0.962r(t-1) - 0.318r(t-2) \\
L(t) &= 10.64 - 0.41t + 0.962r(t-1) - 0.318r(t-2)
\end{aligned}
$$

The solar forecasting models are given by

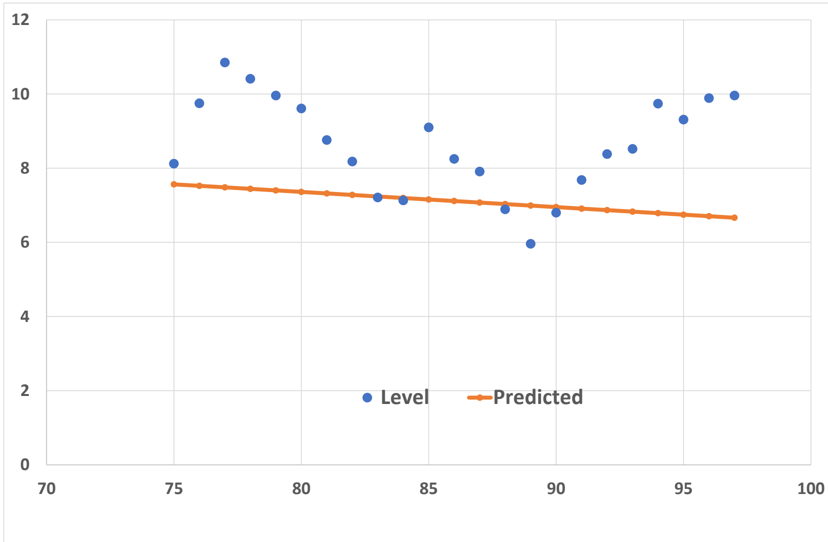

Figure 1: Linear one step ahead forecast

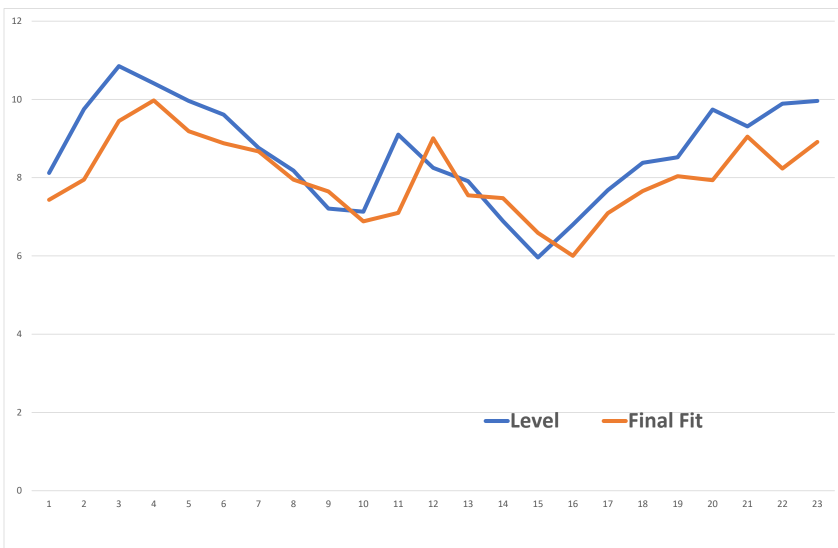

Figure 2: Linear plus AR forecast

$$
\begin{aligned}
f(t) &= 199.9 + 118.17\cos(2\pi t/8760) - 12.06\sin(2\pi t/8760) \\
&+ -90.88\cos(364\pi t/8760) + 8.32\sin(364\pi t/8760) \\
&- 308.28\cos(365\pi t/8760) + 53.95\sin(365\pi t/8760) \\
&- 7.44\cos(366\pi t/8760) + 21.14\sin(366\pi t/8760) \\
&+ 128.96\cos(730\pi t/8760) - 43.89\sin(730\pi t/8760) \\
r(t) &= 0.924r(t-1) - 0.152r(t-2)
\end{aligned}
$$

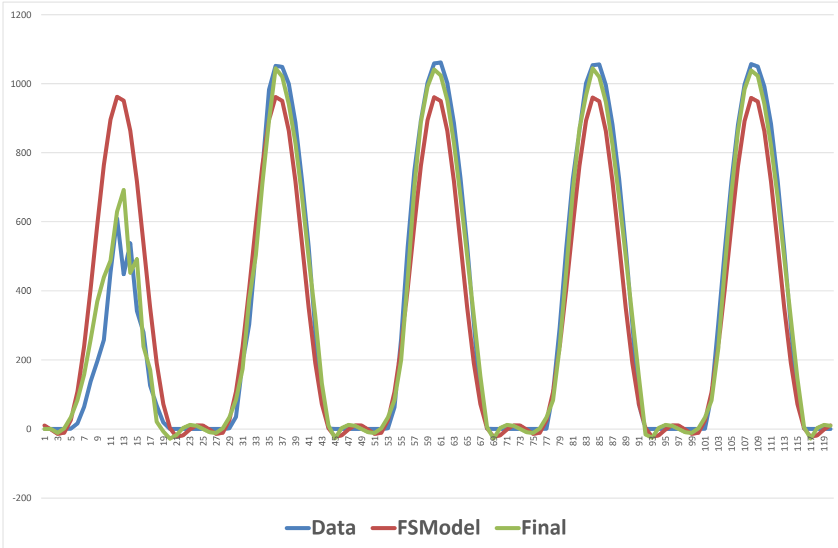

Figure 3: Summer Solar Forecast

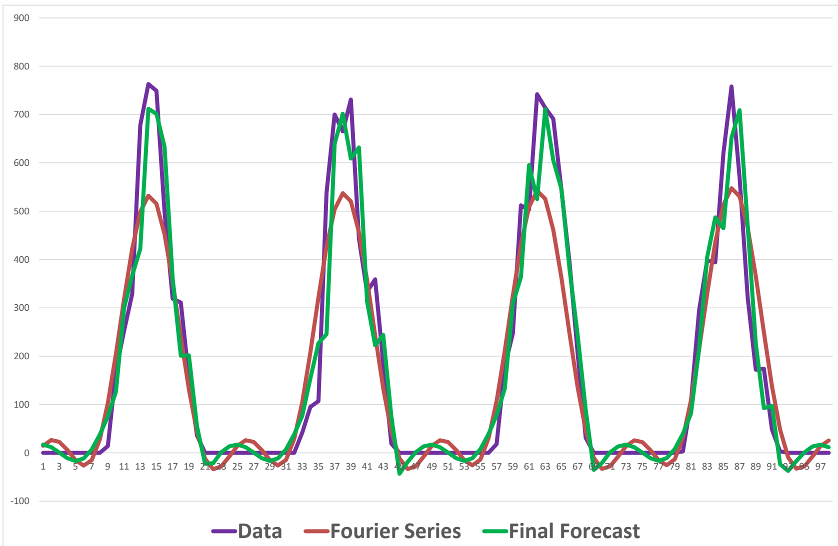

Figure 4: Winter Solar Forecast

3 Error analysis

Error analysis is a very important step for data analysis and forecasting. It is a tool for distinguishing which model is better. In Hoff et al.'s paper, error analyses are classified into two classes. One class consists of absolute dispersion errors which are root mean square error (RMSE) and mean absolute error (MAE). The other class represents relative percentage errors. Relative percentage errors use absolute dispersion error, such as RMSE or MAE divided by the mean of the real data to produce a normalized value. In this section several of the most commonly used error value measures will be introduced. These are median absolute percentage error (MeAPE), mean bias error (MBE), normalised root mean square error (NRMSE), normalised mean absolute error (NMAE), and Kolmogorov-Smirnov test integral (KSI). MeAPE captures the size of the errors and avoids distorting the results for solar and wind energy forecasting. If one uses mean absolute percentage error (MAPE), then with the variables under consideration, wind and solar, there are certain points in time that can produce large errors, and the distribution of errors becomes skewed. So, using MeAPE instead of MAPE for renewable energy forecasting can provide a more accurate perspective. In turn, MBE is used to determine whether any particular model is more biased than another. NRMSE measures overall model quality related to the regression fit. What this means is that is how far the data deviates from the model. What is more informative is, in essence, how far the regression line is from the line $Y = X$, where the $y's$ are the predicted values from the model, and $x's$ are the data values. Interestingly, Willmott and Matsuura produce convincing arguments as to why the mean absolute error (MAE) is a superior error measure to the RMSE. They argue that the RMSE is a function of three characteristics of a set of errors.

It varies with the variability within the distribution of error magnitudes and with the square root of the number of errors ($n^{1/2}$), as well as with the average-error magnitude (MAE).

KSI is a new model validation measure based on the Kolmogorov-Smirnov (KS) test which has the advantage of being non-parametric. KS test is a nonparametric test for the equality of continuous, one dimensional probability distributions. It can be used to compare a sample with a reference probability distribution, in which case it is named a one sample KS test, or to compare two samples, when it becomes a two sample KS test. The KSI measure was proposed in Espinar et al. to assess the similarity of the cumulative distribution functions (CDFs) of actual and modelled data over the whole range of observed values.

Definitions of all the measures are as follows:

1. Median Absolute Percentage Error

$$MeAPE = MEDIAN\left(\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\right) \tag{1}$$

2. Mean Bias Error

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{2}$$

3. Normalised Root Mean Squared Error

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{\bar{y}} \tag{3}$$

where $\hat{y}_i$ are predicted values, y_i are measured values and $\bar{y}$ is the average of measured values.

There is an associated measure, called the Skill Score (SS). It is defined as

$$SS = 1 - \frac{RMSE_{Forecast}}{RMSE_{Reference}} \tag{4}$$

where $RMSE_{Reference}$ is the root mean squared error of some reference or benchmark model. This is usually one of two formulations, the naive persistence forecast, $X_{t+1} = X_t$, or the so-called smart persistence forecast. This is used when the variable under consideration has trend or seasonality embedded in it. An example is the case of hourly solar radiation. In the literature what is usually done is to model the seasonality by defining the clear sky index (CSI). A precursor to this is a clear sky model, a physical model that attempts to give radiation values for a perfectly clear sky for a particular location, for each hour of the year. The CSI is then obtained by dividing the global horizontal radiation (GHI) measured for that hour by the clear sky model for that hour. It is a vexed measure, since Ineichen (2016), a respected researcher in the field, feels the need to check the validity of models - in this paper seven models often used are tested, and three suggested as good. Also, the smart persistence forecast will then include a seasonality model so the results for the Skill Score could well depend on the model chosen.

4. Normalised Mean Absolute Error

$$NMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}(|\hat{y}_i - y_i|)}{\bar{y}} \tag{5}$$

5. Kolmogorov-Smirnov Integral

$$KSI(\%) = 100 \times \frac{\int_{x_{min}}^{x_{max}} D_n dx}{\alpha_{critical}} \tag{6}$$

where x_{max} and x_{min} are the extreme values of the independent variable, and $\alpha_{critical}$ is calculated as $\alpha_{critical} = V_c \times (x_{max} - x_{min})$. The critical value V_c depends on population size N and is calculated for a 99% level of confidence as $V_c = 1.63/\sqrt{N}, N \geq 35$. The D_n are the differences between the cumulative distribution functions (CDF) for each interval. The higher the KSI value, the worse the fit of the model to data.

It is worth noting that in assessing a model against the actual data, NRMSE measures how close the points are clustered around the regression line for the relationship between the observed and predicted values, while KSI and MBE assess the distribution of points around the unit line, $\hat{y} = y$. Thus by considering a set of diverse measures the aim is to allow for a more complete comparison of the proposed models. For example, additional information on the CDFs carried by KSI and MBE can be used to distinguish between models with similar MeAPE or NRMSE values.

4 Error Measures for Lake Huron and Solar

4.1 Lake Huron

Errors	*Linear*	*Final*
NMBE	17.5%	6.9%
NMAE	18.9%	9.3%
NRMSE	22.8%	17.5%

4.2 Solar

Errors	*Linear*	*Final*
NMBE	2.5%	0.21%
NMAE	26.7%	14.5%
NRMSE	33.8%	19.9%

5 Multi-Step Forecasting

Consider a realisation of a stationary series $\{x_i\}$ arising from an ARMA process. Let $\hat{x}_i(l)$ be the *forecast* of x_{i+l} made at the time i. If the length of the observed series to date is large, the estimation errors in the parameters should not generally be large, so the minimum mean square error forecasts can be obtained in the following manner.

We will need the random shock, or moving average, form of the model. For example, let's derive the moving average form of the $AR(1)$ model.

$$
\begin{aligned}
x_t &= \alpha x_{t-1} + z_t \\
&= \alpha(\alpha x_{t-2} + z_{t-1}) + z_t \\
&= \alpha^2 x_{t-2} + \alpha z_{t-1} + z_t \\
&= \alpha^2(\alpha x_{t-3} + z_{t-2}) + \alpha z_{t-1} + z_t \\
&\quad \dots \\
&= \sum_{j=0}^{\infty} \alpha^j z_{t-j}
\end{aligned}
$$

Suppose the best possible forecast is expressible as

$$\hat{x}_i(l) = \psi_l^* z_i + \psi_{l+1}^* z_{i-1} + \dots, \tag{7}$$

where the ψ^* weights are to be determined. The random shock form of the model is

$$x_{i+l} = (z_{i+l} + \psi_1 z_{i+l-1} + \ldots + \psi_{l-1} z_{i+1}) + (\psi_l z_i + \ldots . \tag{8}$$

Thus, the mean square error in the prediction is

$$E[\{x_{i+l} - \hat{x}_i\}^2] = (1 + \psi_1^2 + \ldots + \psi_{l-1}^2)\sigma_z^2 + \sum_{j=l}^{\infty}(\psi_j - \psi_j^*)^2 \sigma_z^2 \tag{9}$$

and this is obviously a minimum when

$$\psi_j^* = \psi_j, \quad j = l, l+1, \ldots .$$

Therefore, the best forecast is

$$\hat{x}_i(l) = \psi_l z_i + \psi_{l+1} z_{i-1} + \ldots . \tag{10}$$

For an $AR(1)$ process for example, the best forecast is

$$\hat{x}_i(l) = \alpha^l z_i + \alpha^{l+1} z_{i-1} + \alpha^{l+2} z_{i-2} + \ldots \tag{11}$$

The *forecast error* at time i for lead l is

$$e_i(l) = z_{i+l} - \hat{x}_i(l) = z_{i+l} + \psi_1 z_{i+l-1} + \ldots + \psi_{l-1} z_{i+1} \tag{12}$$

Let E_i be the *conditional expectation operator*, denoting the 'expected value at time i of'. Then

$$\begin{aligned} E_i[z_{i+l}] &= 0, \quad l > 0 \\ &= a_{i+l}, \quad l \leq 0 \end{aligned} \tag{13}$$

So equation (10) gives

$$E_i[\hat{x}_i(l)] = \hat{x}_i(l), \ \ l > 0 \tag{14}$$

and equation (12) yields

$$E_i[e_i(l)] = 0, \quad l > 0 \tag{15}$$

$$E_i[x_{i+l}] = \hat{x}_i(l), \quad l > 0 \tag{16}$$

$$E_i[x_{i+l}] = x_{i+l}, \quad l < 0 \tag{17}$$

Equation (15) shows that the forecast error has zero expected value and equation (16) shows that the forecasts are unbiased. However, if the forecast function wanders off target to either side, it is likely to remain there in the short run, and the forecast errors at various leads will be correlated. The one-step ahead forecast errors are in fact the shocks which generate the process, and of course are uncorrelated.

From equation (12), the variance of the forecast error at lead l is

$$V(l) = Var[e_i(l)] = (1 + \psi_1^2 + \ldots + \psi_{l-1}^2)\sigma_z^2 \leq \sigma_x^2 \tag{18}$$

and hence the $(100 - \alpha)\%$ confidence intervals for the forecasts are

$$\hat{x}_i(l) \pm k_\alpha \left\{ \left(\sum_{j=0}^{l-1} \psi_j^2 \right) \sigma_z^2 \right\}^{\frac{1}{2}}$$

where k_α is the $(100 - \alpha/2)\%$ point of the standard normal distribution.

6 Example using an $AR(3)$ process

If the forecasting is performed up to l steps ahead, the standard error is calculated up to l steps ahead. The values of ψ_j should be derived to calculate the confidence bounds of the forecast. To do this, the x_i is expressed as a moving average process:

$$\begin{aligned} \phi(B)x_t &= \theta(B)a_t & (19) \\ x_t &= \frac{\theta(B)}{\phi(B)}a_t & (20) \end{aligned}$$

Then

$$x_t = \psi(B)a_t \tag{21}$$

where $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \ldots)$. In this way,

$$\psi(B) = \frac{\theta(B)}{\phi(B)} \tag{22}$$

.

And,

$$
\begin{aligned}
\phi(B)\psi(B) &= \theta(B) && (23)\\
(1-\phi_1 B - \ldots - \phi_p B^p)(1+\psi_1 B + \psi_2 B^2 + \ldots) &= (1-\theta_1 B - \ldots - \theta_q B^q) && (24)
\end{aligned}
$$

Thus, equating the coefficients of powers of B in Eq. 18, the values of ψ_i can be obtained. For an $AR(3)$ process,

$$
\begin{aligned}
(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1+\psi_1 B + \psi_2 B^2 + \ldots) &= 1 && (25)\\
(\psi_1 - \phi_1)B + (\psi_2 - \psi_1\phi_1 - \phi_2)B^2 + (\psi_3 - \phi_1\psi_2 - \phi_2\psi_1 - \phi_3) + & && (26)\\
+(\psi_4 - \phi_1\psi_3 - \phi_2\psi_2 - \phi_3\psi_1)B^4 &= 0 && (27)
\end{aligned}
$$

Given that the values of ϕ_i and θ_i are known, the values of ψ_i are calculated in the following way:

$$
\begin{aligned}
\psi_0 &= 1\\
\psi_1 &= \phi_1\\
\psi_2 &= \psi_1\phi_1 + \phi_2\\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 + \phi_3\\
&\vdots\\
\psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \phi_3\psi_{j-3} && (28)
\end{aligned}
$$

A recursive formula has thus been derived for forecasting at any horizon h when the original process is an $AR(3)$.

7 Other processes

$ARMA(2,1)$

$$
\begin{aligned}
\psi_0 &= 1\\
\psi_1 &= \phi_1 - \theta_1\\
\psi_2 &= \psi_1\phi_1 + \phi_2\\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1\\
&\vdots\\
\psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} && (29)
\end{aligned}
$$

$ARMA(2,2)$

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \phi_1 - \theta_1 \\
\psi_2 &= \psi_1\phi_1 + \phi_2 - \theta_2 \\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 \\
&\vdots \\
\psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} \qquad (30)
\end{aligned}
$$

8 General ARMA(p,q)

The difference equation form of an $ARMA(p,q)$ process is

$$x_{i+l} = \phi_1 x_{i+l-1} + \ldots + \phi_p x_{i+l-p} + z_{i+l} + \theta_1 z_{i+l-1} + \ldots + \theta_q z_{i+l-q} \qquad (31)$$

So, taking conditional expectations at time i, for $l > 0$ the optimal forecast function is

$$\hat{x}_i(l) = \phi_1 E_i[x_{i+l-1}] + \ldots + \phi_p E_i[x_{i+l-p}] + 0 + \theta_1 E_i[z_{i+l-1}] + \ldots + \theta_q E_i[z_{i+l-q}]. \qquad (32)$$

Thus for $1 < l < p, q$ say,

$$\hat{x}_i(l) = \phi_1 \hat{x}_i(l-1) + \ldots + \phi_{l-1}\hat{x}_i(1) + \phi_l x_i + \ldots + \phi_p x_{i+1-p} + \theta_l z_i + \ldots + \theta_q z_{i+l-q} \qquad (33)$$

An observed series of length N for which forecasts are required for leads up to L, we calculate the $\hat{x}_N(l),\ l = 1, \ldots, L$ recursively to fill up line (A) of Table **??** where if necessary some initial z's may need to be set to zero.

			$leadl$					
i	x_i	z_i	1	2	$\ldots$	$L-1$	L	
N	x_N		$\hat{x}_N(1)$	$\hat{x}_N(2)$	$\ldots$	$\hat{z}_N(L-1)$	$\hat{z})N(L)$	(A)
$N+1$	x_{N+1}	z_{N+1}	$\hat{x}_{N+1}(1)$	$\hat{x}_{N+1}(2)$	$\ldots$	$\hat{x}_{N+1}(L-1)$	$\hat{x}_{N+1}(L)$	(B)
$Conf.Limits$			λ_1	λ_2	$\ldots$	λ_{L-1}	λ_L	(C)
			$where\lambda_l = k_\alpha \left\{ \left(\sum_{j=0}^{l-1} \psi_j^2 \right) \sigma_z^2 \right\}^{\frac{1}{2}},\ \ l = 1, \ldots, L$					

When x_{N+1} becomes available, we obtain

$$z_{N+1} = x_{N+1} - \hat{x}_N(1)$$

and then can update the forecasts for the leads $1, \ldots, L-1$ using

$$\hat{x}_{N+1}(l) = \hat{x}_N(l+1) + \psi_l z_{N+1}$$

This follows since equation (**??**) gives

$$\hat{x}_{N+1}(l) = \psi_l z_{N+1} + \psi_{l+1} z_N + \psi_{l+2} z_{N-1} + \ldots$$

and

$$\hat{x}_N(l+1) = \psi_{l+1} z_N + \psi_{l+2} z_{N-1} + \ldots$$

8.1 Examples

1. Assume we have a series $x_i = z_i - 0.465 z_i$, $\sigma_z^2 = 0.872$. with $x_{97} = -9406$. Find the forecast for 3 leads, and then take into account a value for $x_{98} = 0.49$ to update these forecasts. Also, work out confidence intervals.

   Solution: Equation (**??**) gives an estimate for

   $$\hat{x}_{97}(1) = -0.465 \times -0.9406 = 0.4374$$

   and

   $$\hat{x}_{97}(2) = \hat{x}_{98} = 0$$

   To calculate the confidence intervals we need

   $$\begin{aligned} \lambda_1 &\simeq 2\sqrt{1 \times 0.872} = 1.9 \\ \lambda_2 &\simeq 2\sqrt{(1 + 0.465^2) \times 0.872} = 2.1 \end{aligned}$$

   This means we have confidence intervals of $(-1.5, 2.3)$ and $(-2.1, 2.1)$ for x_{98} and x_{99} respectively. Note that this is as far as we can go in calculating confidence intervals since we only have an $MA(1)$ process.

9 Alternative method for multi-step forecasting

I shall just demonstrate with an $AR(3)$ model

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3} + z_t. \tag{34}$$

The one step ahead forecast at time t is written as

$$\hat{x}_{t+1} = \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 x_{t-2}. \tag{35}$$

If one wants to forecast two steps ahead at time t, then one does not have a value for x_{t+1} but there is a forecast of that value, and that is what we will use. Thus, for two steps ahead at time t,

$$\hat{x}_{t+2} = \alpha_1 \hat{x}_{t+1} + \alpha_2 x_t + \alpha_3 x_{t-1}. \tag{36}$$

This process can continue on but of course after one more step, there are no more observed values to use, so the forecast becomes dependent only on the forecasted values. This is altered if there also was a trend or seasonality when those components are added to the forecast.

10 Forecasting a GARCH(1,1) process

The $GARCH(1,1)$ process is characterised by

$$\begin{aligned}
a_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 a_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\
&= \alpha_0(1+\beta_1) + \alpha_1(a_{t-1}^2 + \beta_1 a_{t-2}^2) + \beta_1^2 \sigma_{t-2}^2 \\
&= \alpha_0(1+\beta_1) + \alpha_1(a_{t-1}^2 + \beta_1 a_{t-2}^2) + \beta_1^2(\alpha_0 + \alpha_1 a_{t-3}^2 + \beta_1 \sigma_{t-3}^2) \\
&= \alpha_0(1+\beta_1+\beta_1^2) + \alpha_1(a_{t-1}^2 + \beta_1 a_{t-2}^2 + \beta_1^2 a_{t-3}^2) + \beta_1^3 \sigma_{t-3}^2 \\
&= \alpha_0(1+\beta_1+\beta_1^2+\ldots) + \alpha_1(a_{t-1}^2 + \beta_1 a_{t-2}^2 + \beta_1^2 a_{t-3}^2 + \ldots) \\
&= \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} a_{t-j}^2
\end{aligned}$$

We use this relationship to forecast into the future.