Linear prediction

Andreas Jakobsson

LUND UNIVERSITY

Linear prediction

”*It is not hard to make predictions. It is just tough to make ones that turn out to be right.*”

John Treichler

”*It is tough to make predictions, especially about the future.*”

Yogi Berra

Linear prediction

A *linear* prediction is formed as a weighted sum of earlier observations, such that

$$\hat{y}_{t+k|t} = \sum_{\ell=0}^{n} w_\ell y_{t-\ell}$$

For a Gaussian process, the optimal predictor is linear, and is thus the same as the optimal linear predictor. This predictor is determined such that

$$C\Big\{y_{N+k} - \hat{y}_{N+k|N}, y_t\Big\} = C\Big\{y_{N+k} - \sum_{\ell=0}^{n} w_\ell y_{N-\ell}, y_t\Big\} = 0$$

That is, the optimal predictor is such that the resulting prediction error is uncorrelated with the earlier observed measurements. This is but a reformulation of the orthogonality principle.

Linear prediction

A *linear* prediction is formed as a weighted sum of earlier observations, such that

$$\hat{y}_{t+k|t} = \sum_{\ell=0}^{n} w_\ell y_{t-\ell}$$

For a Gaussian process, the optimal predictor is linear, and is thus the same as the optimal linear predictor. This predictor is determined such that

$$C\Big\{y_{N+k} - \hat{y}_{N+k|N}, y_t\Big\} = C\Big\{y_{N+k} - \sum_{\ell=0}^{n} w_\ell y_{N-\ell}, y_t\Big\} = 0$$

That is, the optimal predictor is such that the resulting prediction error is uncorrelated with the earlier observed measurements. This is but a reformulation of the orthogonality principle.

The optimal (linear) k-step prediction error is

$$\begin{aligned}\mathcal{C} &= E\left\{\left[y_{N+k} - \hat{y}_{N+k|N}\right]^2\right\} = V\left\{y_{N+k} - \hat{y}_{N+k|N}\right\} \\ &= C\left\{y_{N+k} - \hat{y}_{N+k|N}, y_{N+k}\right\} \\ &= V\left\{y_{N+k}\right\} - \sum_{\ell=0}^{n} w_\ell C\left\{y_{N-\ell}, y_{N+k}\right\}\end{aligned}$$

as $y_{N+k} - \hat{y}_{N+k|N}$ is orthogonal to $\hat{y}_{N+k|N}$.

Linear prediction

Example: Consider the MA process $y_t = e_t + c_1 e_{t-1}$, and assume that y_1 and y_2 have been observed. The optimal linear prediction of y_3 can then be found by noting that

$$C\{y_3 - w_1 y_1 - w_2 y_2, y_k\} = 0$$

for $k = 1$ and $k = 2$. Thus,

$$\begin{aligned} r_y(2) - w_1 r_y(0) + w_2 r_y(1) &= 0 \\ r_y(1) - w_1 r_y(1) + w_2 r_y(0) &= 0 \end{aligned}$$

implying that

$$r_y(0) = (1 + c_1^2)\sigma_e^2 \quad r_y(1) = c_1 \sigma_e^2 \quad r_y(2) = 0$$

and thus

$$w_1 = -\frac{\rho_y^2(1)}{1 - \rho_y^2(1)} \quad \text{and} \quad w_2 = -\frac{\rho_y(1)}{1 - \rho_y^2(1)}$$

where $\rho_y(1) = c_1/(1 + c_1^2)$. The predictor is therefore given as

$$\hat{y}_{3|y_1,y_2} = -\frac{c_1^2}{1 + c_1^2 + c_1^4} y_1 + \frac{c_1 + c_1^3}{1 + c_1^2 + c_1^4} y_2$$

Prediction of ARMA processes

We will now consider predicting an ARMA(p,q), such that

$$A(z)y_t = C(z)e_t$$

implying that

$$y_{t+k} = \frac{C(z)}{A(z)}e_{t+k} = \sum_{\ell=0}^{\infty} \psi_\ell e_{t+k-\ell}$$

Prediction of ARMA processes

We will now consider predicting an ARMA(p,q), such that

$$A(z)y_t = C(z)e_t$$

implying that

$$\begin{aligned} y_{t+k} &= \frac{C(z)}{A(z)}e_{t+k} = \sum_{\ell=0}^{\infty} \psi_\ell e_{t+k-\ell} \\ &= \sum_{\ell=0}^{k-1} \psi_\ell e_{t+k-\ell} + \sum_{\ell=k}^{\infty} \psi_\ell e_{t+k-\ell} \\ &= F(z)e_{t+k} + \sum_{\ell=k}^{\infty} \psi_\ell e_{t+k-\ell} \end{aligned}$$

Prediction of ARMA processes

We will now consider predicting an ARMA(p,q), such that

$$A(z)y_t = C(z)e_t$$

implying that

$$\begin{aligned} y_{t+k} &= \frac{C(z)}{A(z)}e_{t+k} = \sum_{\ell=0}^{\infty}\psi_\ell e_{t+k-\ell} \\ &= \sum_{\ell=0}^{k-1}\psi_\ell e_{t+k-\ell} + \sum_{\ell=k}^{\infty}\psi_\ell e_{t+k-\ell} \\ &= F(z)e_{t+k} + \sum_{\ell=k}^{\infty}\psi_\ell e_{t+k-\ell} \\ &= F(z)e_{t+k} + \sum_{\ell=0}^{\infty}\psi_\ell e_{t-\ell} \end{aligned}$$

where $F(z)$ is monic as $A(z)$ and $C(z)$ are. Thus,

$$F(z) = 1 + f_1 z^{-1} + \ldots + f_{k-1}z^{-k+1}$$

Prediction of ARMA processes

Proceeding, let

$$
\begin{aligned}
y_{t+k} &= F(z)e_{t+k} + \sum_{\ell=0}^{\infty} \psi_\ell e_{t-\ell} \\
&= F(z)e_{t+k} + \frac{G(z)}{A(z)} e_t \\
&= F(z)e_{t+k} + z^{-k} \frac{G(z)}{A(z)} e_{t+k}
\end{aligned}
$$

where the polynomials $G(z)$ and $F(z)$ satisfy the Diophantine equation

$$
C(z) = A(z)F(z) + z^{-k}G(z)
$$

with

$$
\begin{aligned}
\operatorname{ord}\{F(z)\} &= k - 1 \\
\operatorname{ord}\{G(z)\} &= \max(p - 1,\, q - k)
\end{aligned}
$$

with $\operatorname{ord}\{F(z)\}$ denoting the order of the polynomial $F(z)$. Note that $G(z)$ is generally not monic.

Prediction of ARMA processes

The optimal linear prediction is formed as

$$\hat{y}_{t+k|t}(\boldsymbol{\Theta}) = E\left\{y_{t+k}|\boldsymbol{\Theta}\right\}$$

where

$$\boldsymbol{\Theta} = \left[\begin{array}{cc} \boldsymbol{\theta}^T & \mathbf{Y}_t^T \end{array}\right]^T$$

with $\boldsymbol{\theta}$ denoting the model parameters and

$$\mathbf{Y}_t = \left[\begin{array}{ccc} y_1 & \dots & y_t \end{array}\right]^T$$

Prediction of ARMA processes

The optimal linear prediction is formed as

$$\hat{y}_{t+k|t}(\mathbf{\Theta}) = E\left\{y_{t+k}|\mathbf{\Theta}\right\}$$

where

$$\mathbf{\Theta} = \left[\begin{array}{cc} \boldsymbol{\theta}^T & \mathbf{Y}_t^T \end{array}\right]^T$$

with $\boldsymbol{\theta}$ denoting the model parameters and

$$\mathbf{Y}_t = \left[\begin{array}{ccc} y_1 & \ldots & y_t \end{array}\right]^T$$

Thus,

$$\begin{aligned}
\hat{y}_{t+k|t}(\mathbf{\Theta}) &= E\left\{y_{t+k}|\mathbf{\Theta}\right\} \\
&= E\left\{F(z)e_{t+k}\middle|\mathbf{\Theta}\right\} + E\left\{z^{-k}\frac{G(z)}{A(z)}e_{t+k}\middle|\mathbf{\Theta}\right\} \\
&= E\left\{z^{-k}\frac{G(z)}{A(z)}e_{t+k}\middle|\mathbf{\Theta}\right\} \\
&= E\left\{\frac{G(z)}{A(z)}\frac{A(z)}{C(z)}y_t\middle|\mathbf{\Theta}\right\} = \frac{G(z)}{C(z)}y_t
\end{aligned}$$

Prediction of ARMA processes

The part that may not be predicted, the prediction error, may thus be written as

$$\begin{aligned}
\epsilon_{t+k|t}(\mathbf{\Theta}) &= y_{t+k} - \hat{y}_{t+k}(\mathbf{\Theta}) \\
&= F(z)e_{t+k} + \frac{G(z)}{C(z)}y_t - \frac{G(z)}{C(z)}y_t \\
&= F(z)e_{t+k}
\end{aligned}$$

implying that the prediction error should behave as an MA$(k-1)$ process, with

$$V\left\{\epsilon_{t+k|t}(\mathbf{\Theta})\right\} = \left(1 + f_1^2 + \ldots + f_{k-1}^2\right)\sigma_e^2$$

For a Normal distributed process, the $(1-\alpha)$ confidence prediction interval can therefore be expressed as

$$\hat{y}_{t+k|t} \pm u_{\alpha/2}\sigma_e\sqrt{1 + f_1^2 + \ldots + f_{k-1}^2}$$

where $u_{\alpha/2}$ denotes the $\alpha/2$ quantile in the standard Normal distribution.

Prediction of ARMA processes

Example: Compute the 5-step prediction of the SARIMA process

$$(1 - 0.2z^{-1})\nabla_{12} y_t = (1 - 0.3z^{-12})e_t$$

Thus,

$$\begin{aligned} A(z) &= (1 - 0.2z^{-1})(1 - z^{-12}) = 1 - 0.2z^{-1} - z^{-12} + 0.2z^{-13} \\ C(z) &= 1 - 0.3z^{-12} \end{aligned}$$

implying that $p = \mathrm{ord}\{A(z)\} = 13$ and $q = \mathrm{ord}\{C(z)\} = 12$. For $k = 5$,

$$\begin{aligned} \mathrm{ord}\{F(z)\} &= 5 - 1 = 4 \\ \mathrm{ord}\{G(z)\} &= \max(13 - 1,\, 12 - 5) = 12 \end{aligned}$$

Performing the polynomial division yields

$$\begin{aligned} F(z) &= 1 + 0.2z^{-1} + 0.2^2 z^{-2} + 0.2^3 z^{-3} + 0.2^4 z^{-4} \\ G(z) &= 0.2^5 + 0.7z^{-7} - 0.2^5 z^{-12} \end{aligned}$$

Thus, $(1 - 0.3z^{-12})\hat{y}_{t+5|t}(\boldsymbol{\Theta}) = (0.2^5 + 0.7z^{-7} - 0.2^5 z^{-12})y_t$, implying that

$$\hat{y}_{t+5|t}(\boldsymbol{\Theta}) = 0.2^5 y_t + y_{t-7} - 0.2^5 y_{t-12}$$

Prediction of ARMA processes

Matlab example of 5-step prediction of

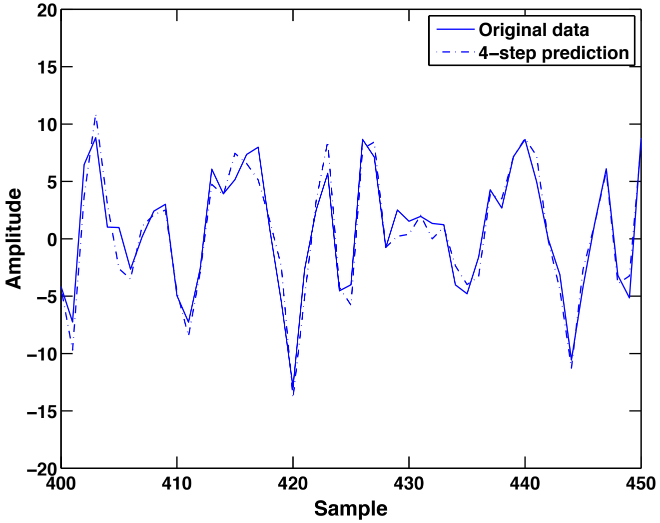
$$(1 - 0.2z^{-1})\nabla_{12} y_t = (1 - 0.3z^{-12})e_t$$

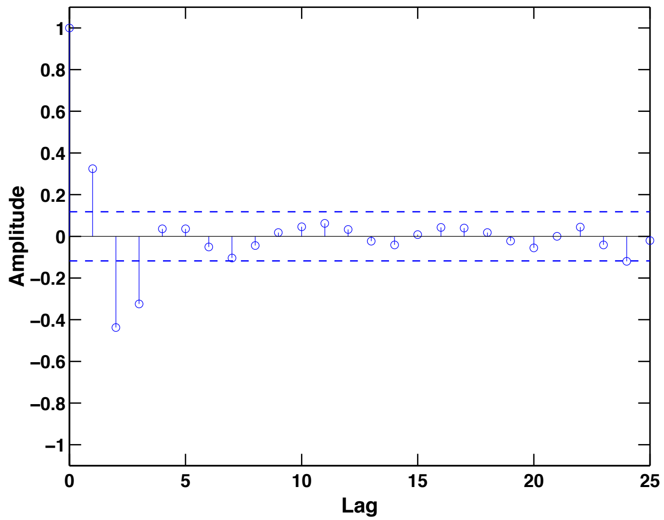
Code:

```
s = 12;
k = 5;
C = [ 1 zeros(1,11) -0.3 ];
A = conv( [ 1 -0.2 ], [ 1 zeros(1,s-1) -1 ] );
[F,G] = polydiv( C, A, k );
```

One of the provided Matlab functions.

Prediction of ARMA processes





Example:
Compute the 4-step prediction of the SARIMA process

$$(1 + 0.8z^{-1} + 0.8z^{-2})\nabla_{24} y_t = (1 + 0.4z^{-1} + 0.6z^{-14})e_t$$

For this process,

$$p = \text{ord}\{A(z)\} = 26$$
$$q = \text{ord}\{C(z)\} = 14$$

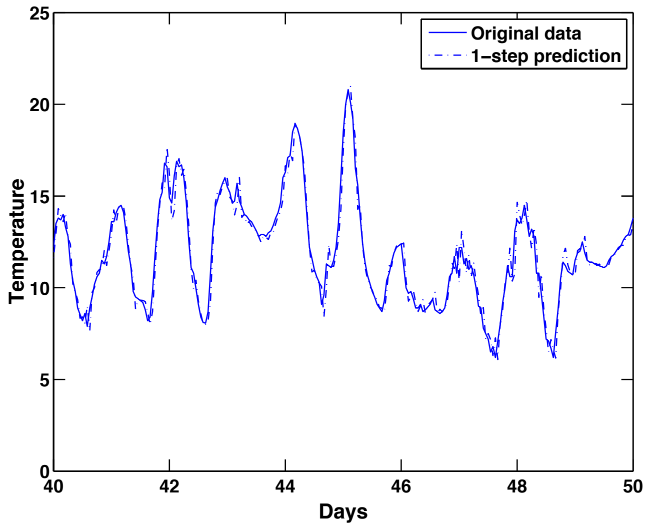
Thus, $\text{ord}\{F(z)\} = k-1 = 3$ and $\text{ord}\{G(z)\} = \max(p-1,\ q-k) = 25$, yielding

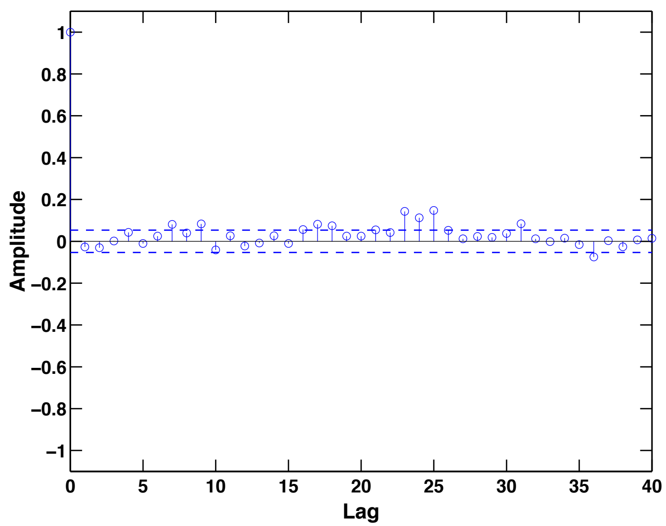
$F(z) = 1 - 0.4z^{-1} - 0.48z^{-2} + 0.704z^{-3}$

$G(z) = -0.1792 - 0.5632z^{-1} + 0.6z^{-10} + z^{-20} + 0.4z^{-21} + 0.1792z^{-24} + 0.5632z^{-25}$

Note that the prediction error behaves like an MA(3).

Prediction of ARMA processes



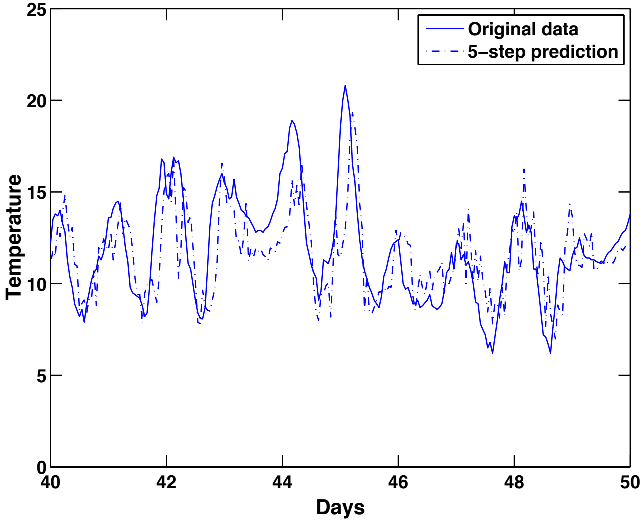


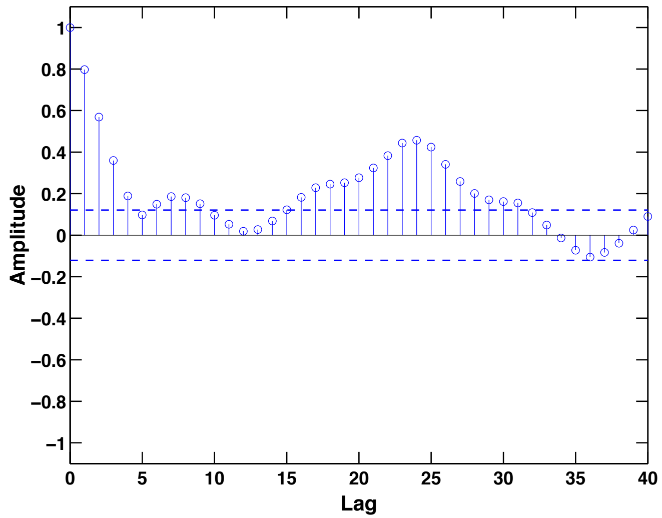
Example:
Consider the following initial model for the temperature in Svedala:

$$A(z) = 1 - 1.79z^{-1} + 0.84z^{-2}$$
$$C(z) = 1 - 0.18z^{-1} - 0.11z^{-2}$$

Forming a one-step prediction using this model - is the prediction residual white?
The ACF is deemed normal distributed.

Prediction of ARMA processes



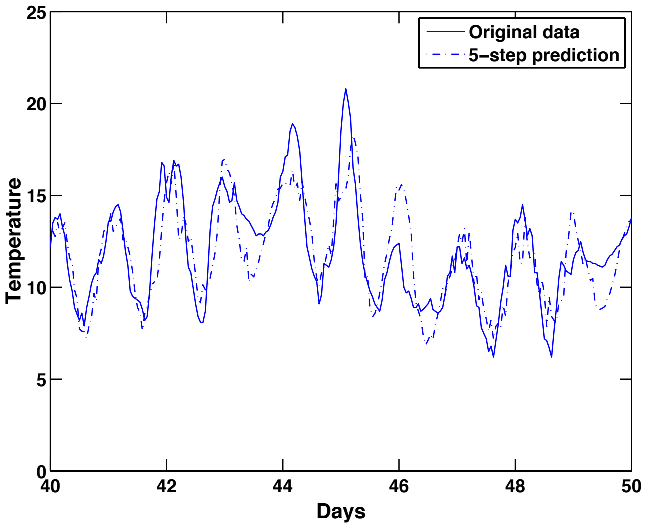


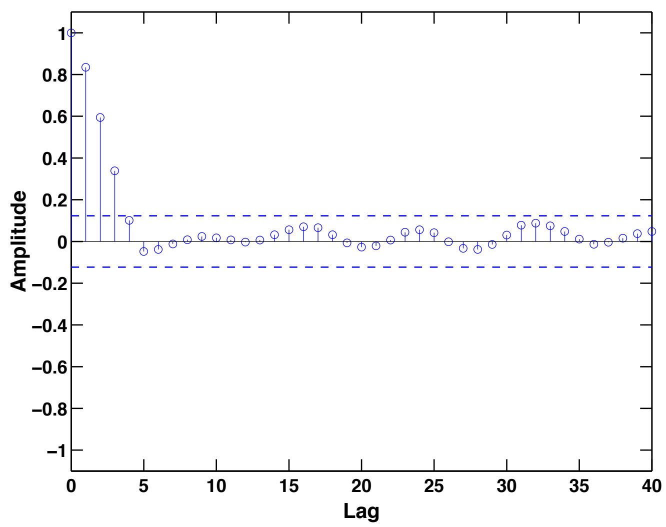
Example:
Consider the following initial model for the temperature in Svedala:

$$A(z) = 1 - 1.79z^{-1} + 0.84z^{-2}$$
$$C(z) = 1 - 0.18z^{-1} - 0.11z^{-2}$$

Proceeding, we form a 5-step prediction. Is the model sufficient?

Prediction of ARMA processes



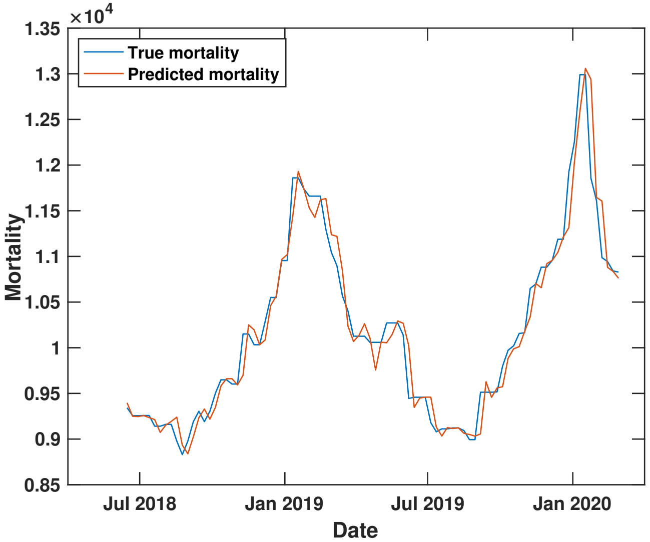


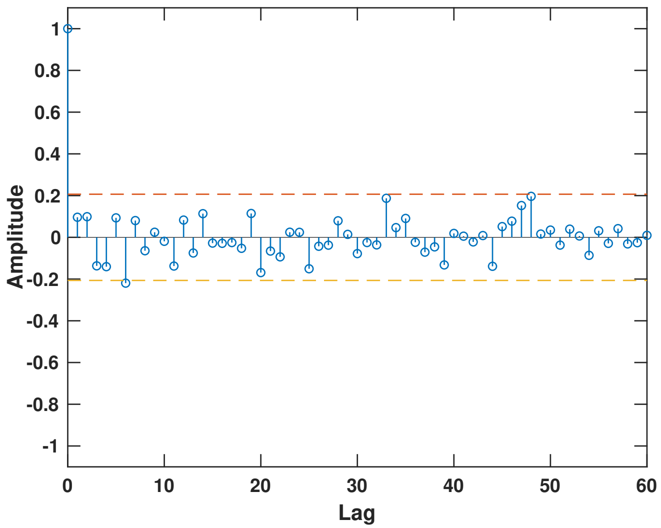
Example:
Extending (and re-estimating) the model, we obtain:

$$A(z) = 1 - 3.7511z^{-1} + 5.3368z^{-2} - 3.4079z^{-3} + 0.8224z^{-4}$$
$$C(z) = 1 - 2.3056z^{-1} + 1.4174z^{-2} + 0.0773z^{-3} - 0.0633z^{-4} - 0.1176z^{-5}$$

Proceeding, we form a 5-step prediction. Is the model sufficient?

Mortality





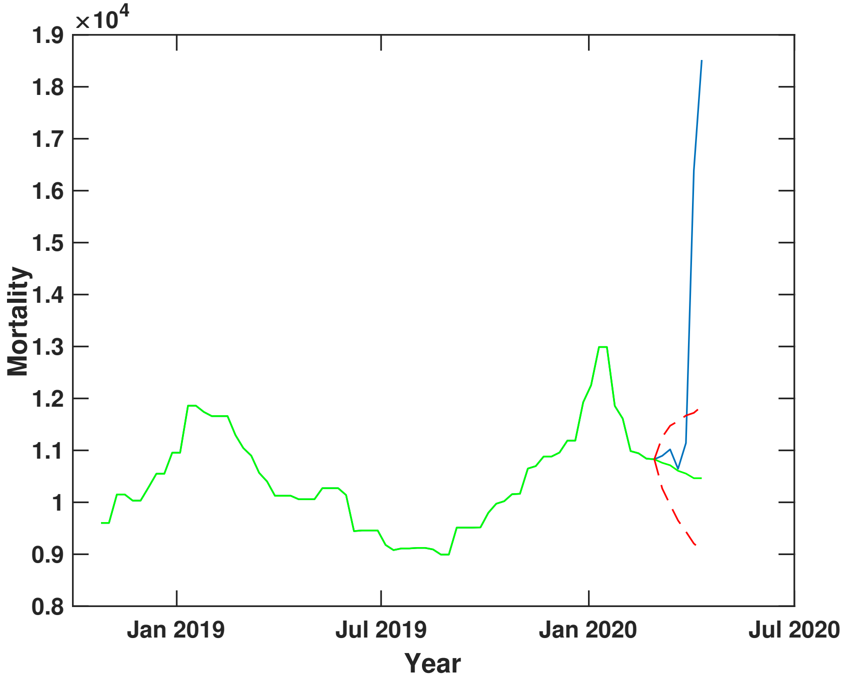
Using the model we earlier constructed for the mortality data in the UK

$$(1 + a_{52}z^{-52})\nabla y_t = (1 + c_1 z^{-1})e_t$$

we proceed to form a one-step prediction of the mortality in our validation data. Note that such data may *not* be used to construct the model.

The resulting prediction residual pass our whiteness tests (and the ACF can be deemed to be Gaussian distributed).

Mortality



Given the model, we can predict the expected mortality at time $t+1, t+2, \ldots$, starting from the beginning of the outbreak of the Corona pandemic.

Note that one needs to recompute $F(z)$ and $G(z)$ for each such prediction, with each step yielding its own prediction variance. This will clearly grow the further one tries to predict ahead.

Comparing the prediction with the actual outcome, one may conclude that the pandemic during these few weeks caused about 15.000 additional fatalities.