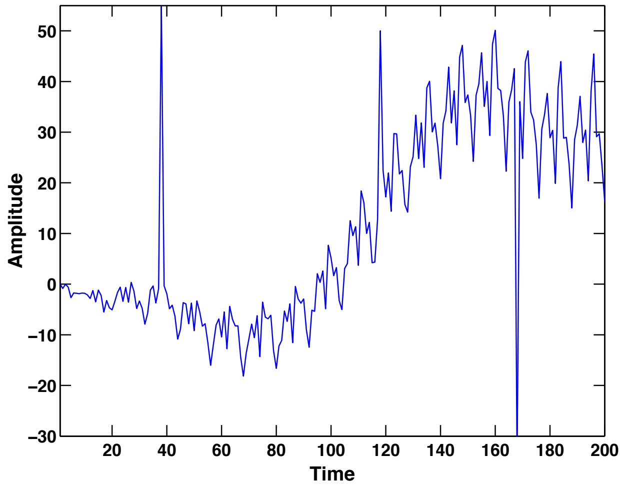
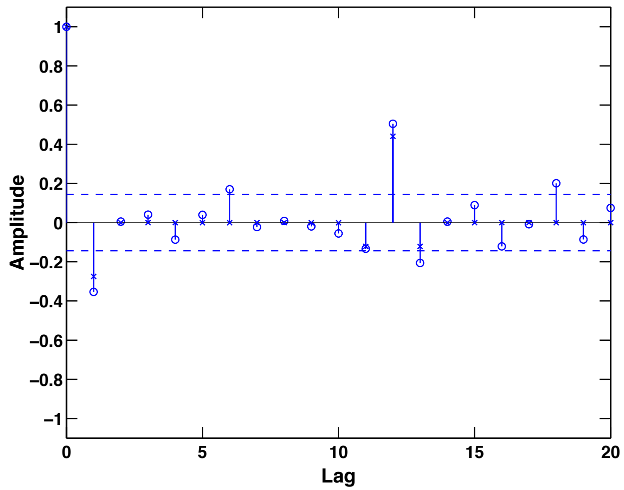
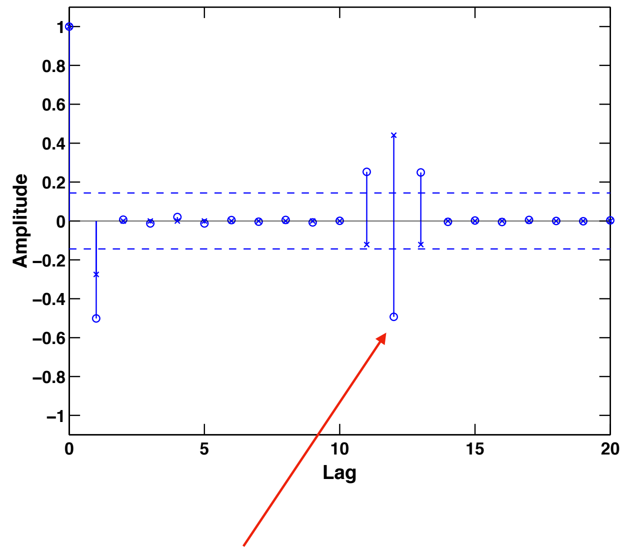
Outliers and irregularities

Andreas Jakobsson

Outliers and irregularities







Three (fairly obvious) outliers have been added to the process

$$\nabla\nabla_{12}y_t = (1 - 0.3z^{-1})(1 + 0.5z^{-12})$$

These will need to handled as they will corrupt the resulting estimates.

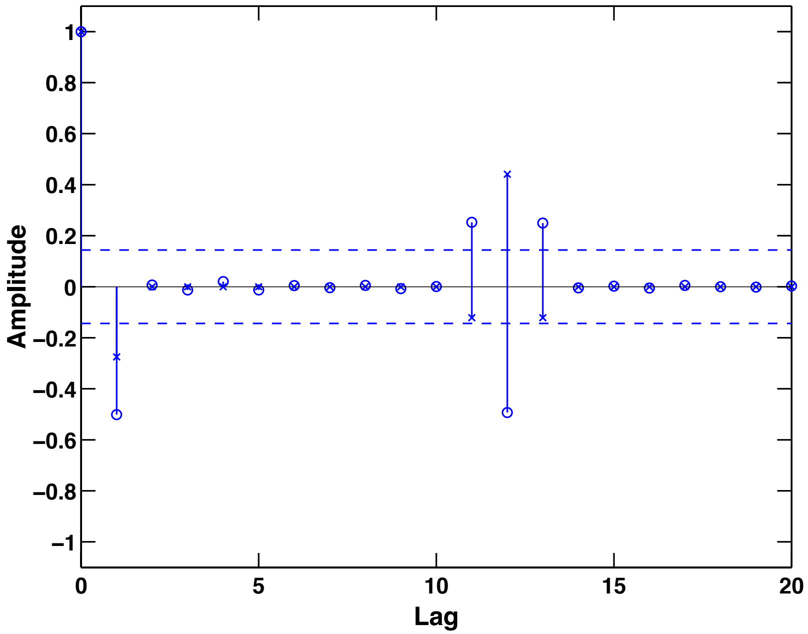
The figure shows the estimated ACF as compared to true values (marked with crosses). The left figures did not contain the outliers, the right one did.

This is mainly done by

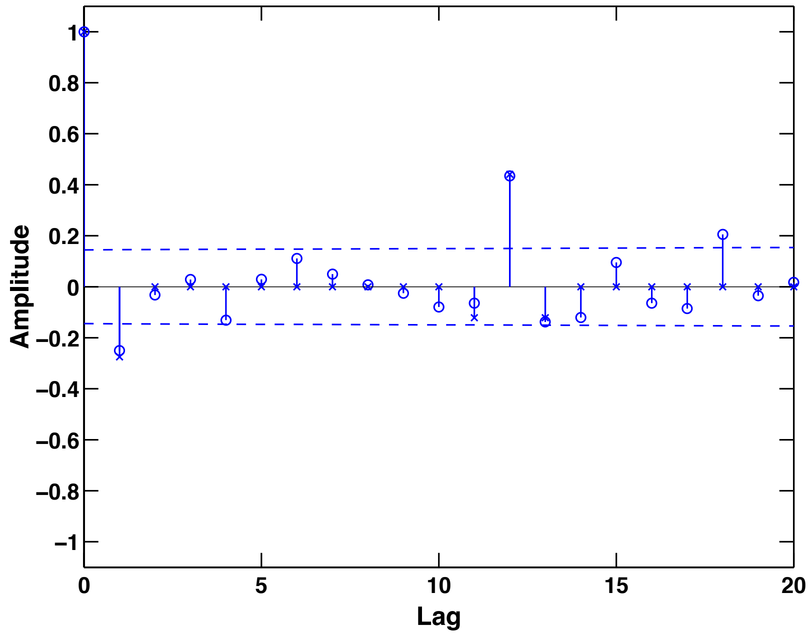
(i) detecting the outliers and remove them, possibly replacing the missing samples with interpolated data, or

(ii) using robust estimators that can handle the presence of outliers.

The trimmed autocorrelation function



ACF estimate

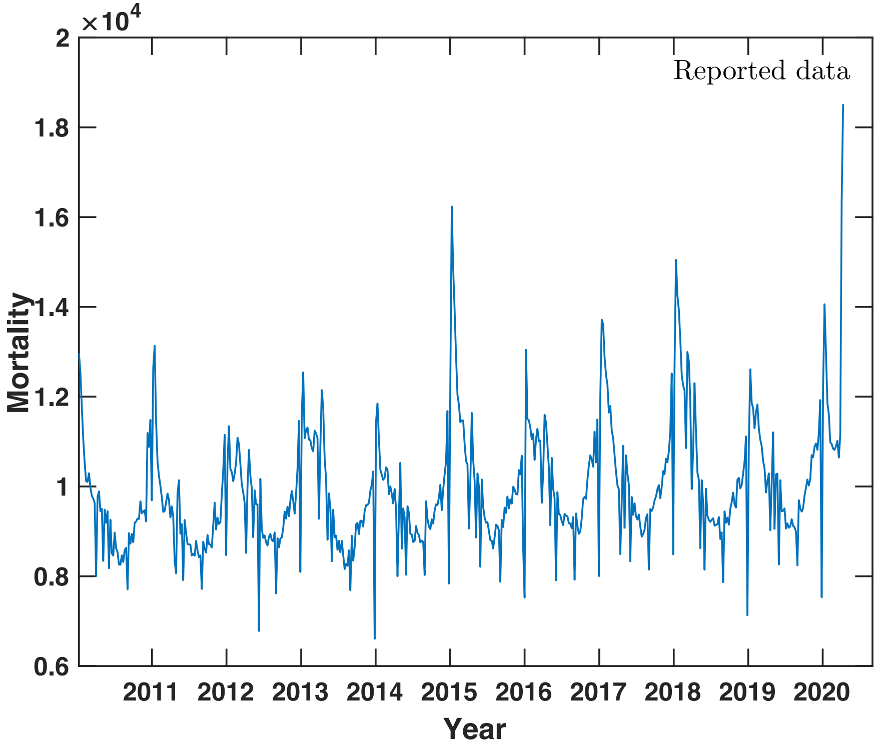


TACF estimate

A simple way to omit outliers, that often works well, is to sort the data and simply remove the tail values, before resorting the data points. The removed samples are then missing, so a bit of bookkeeping is required when forming the estimate.

The α-trimmed autocorrelation estimate removes the α largest and smallest values before computing the ACF. Use the provided function `tacf`.

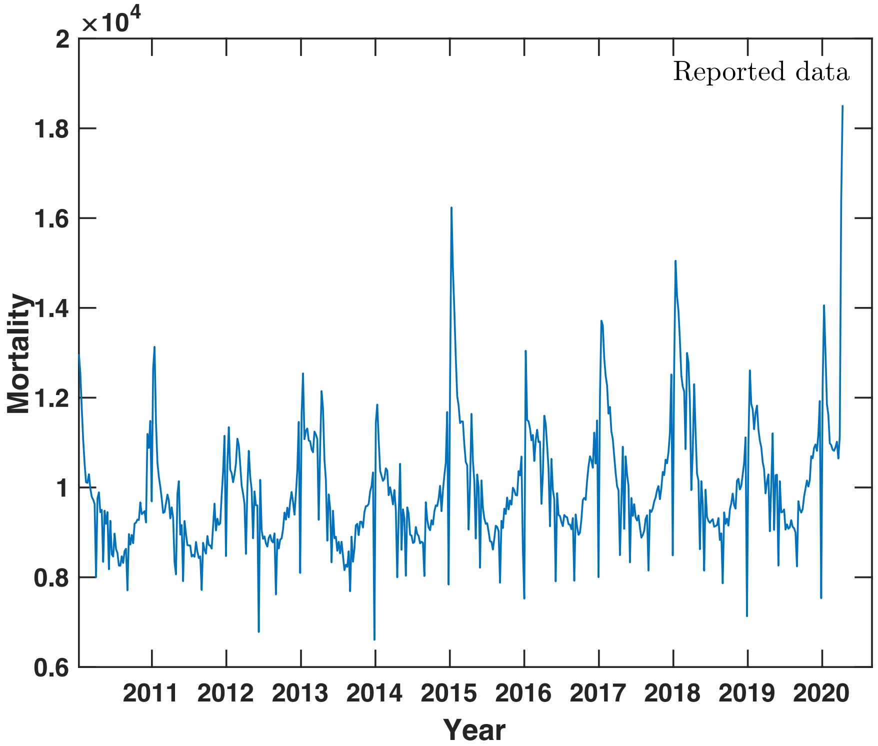
Mortality data

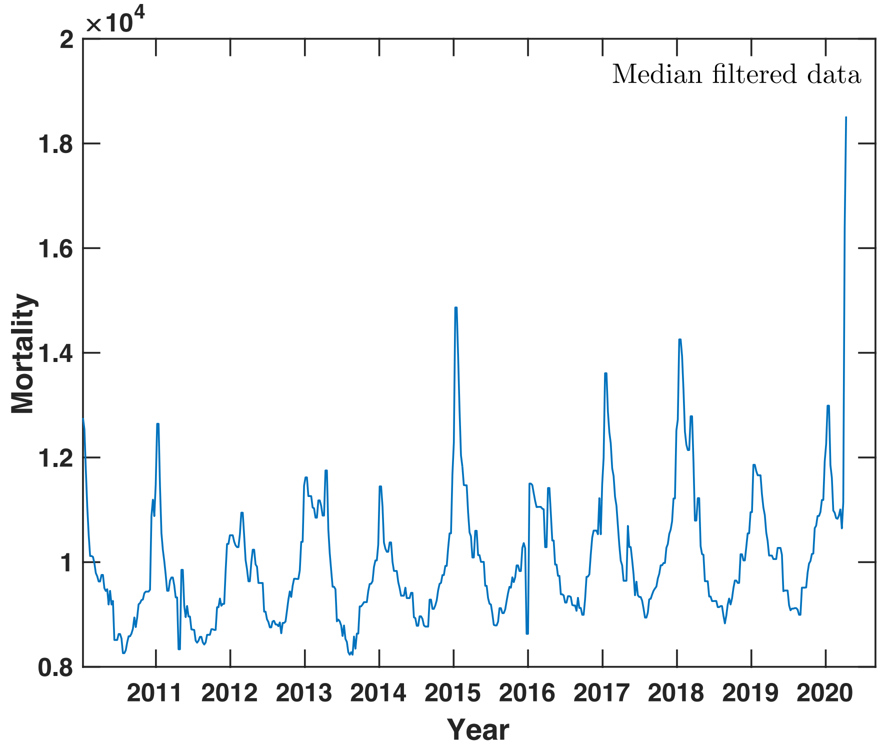


The weekly mortality data in the United Kingdom. The onset of the Corona pandemic is clearly visible in early 2020.

Due to the way deaths are reported, there are "jump irregularities" in connection with, for instance, Christmas.

Mortality data



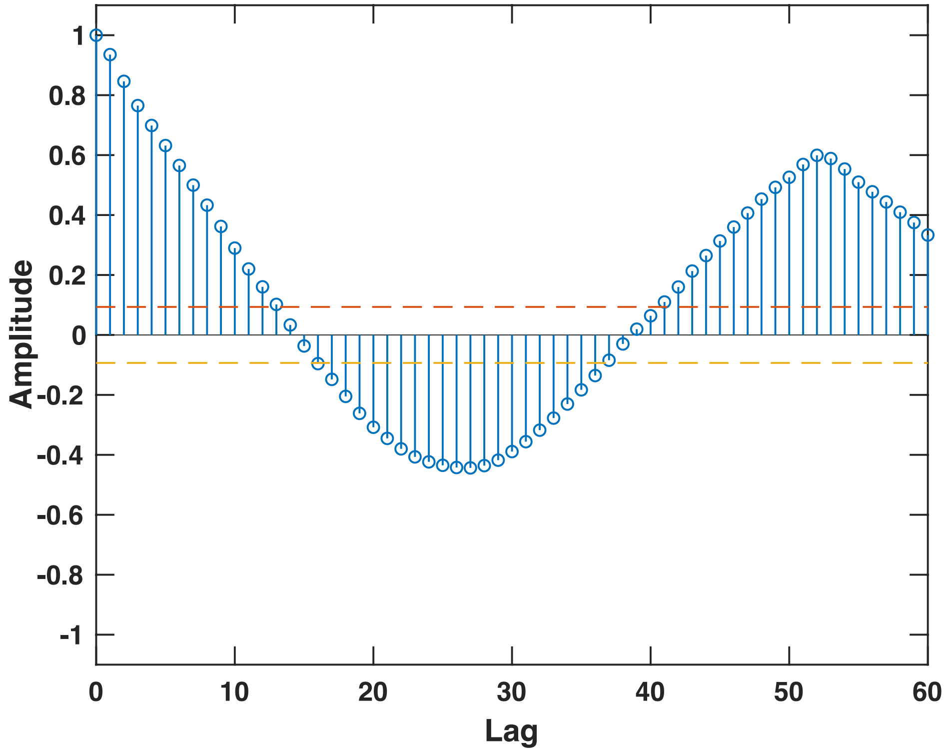


The weekly mortality data in the United Kingdom. The onset of the Corona pandemic is clearly visible in early 2020.

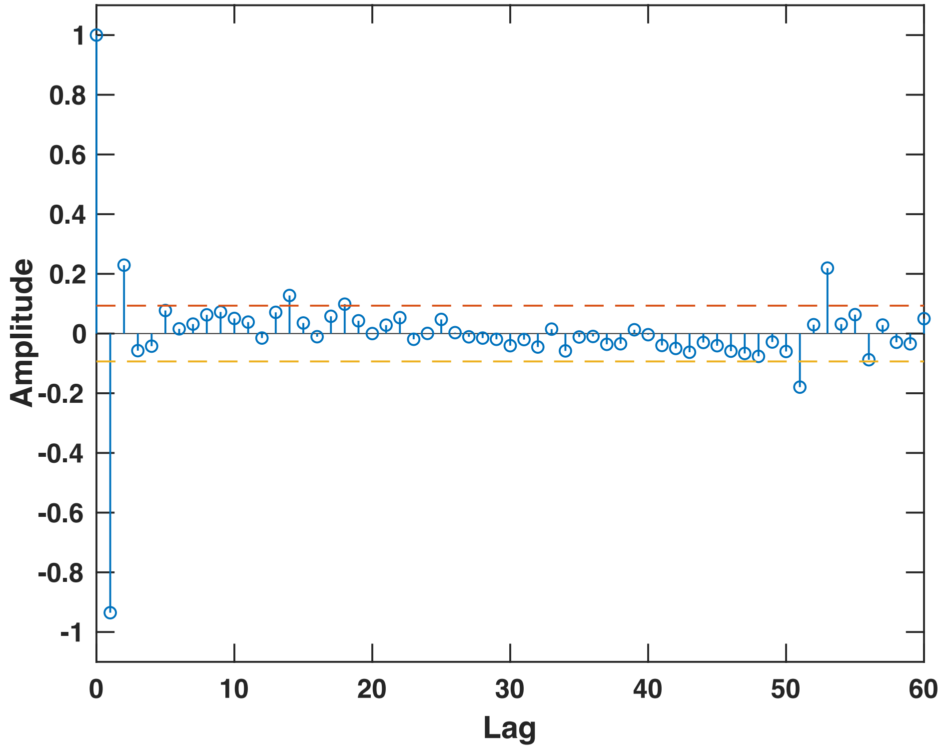
Due to the way deaths are reported, there are ”jump irregularities” in connection with, for instance, Christmas.

Note that the data has a growing trend, suggesting that we might need to differentiate the data.

Mortality data



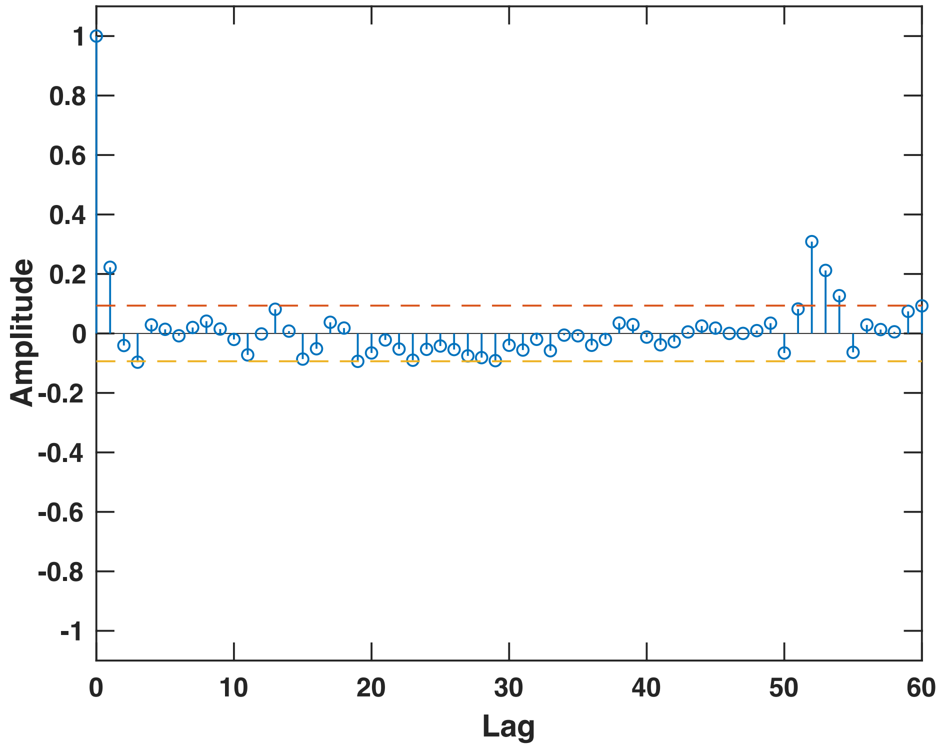
ACF of ∇y_t



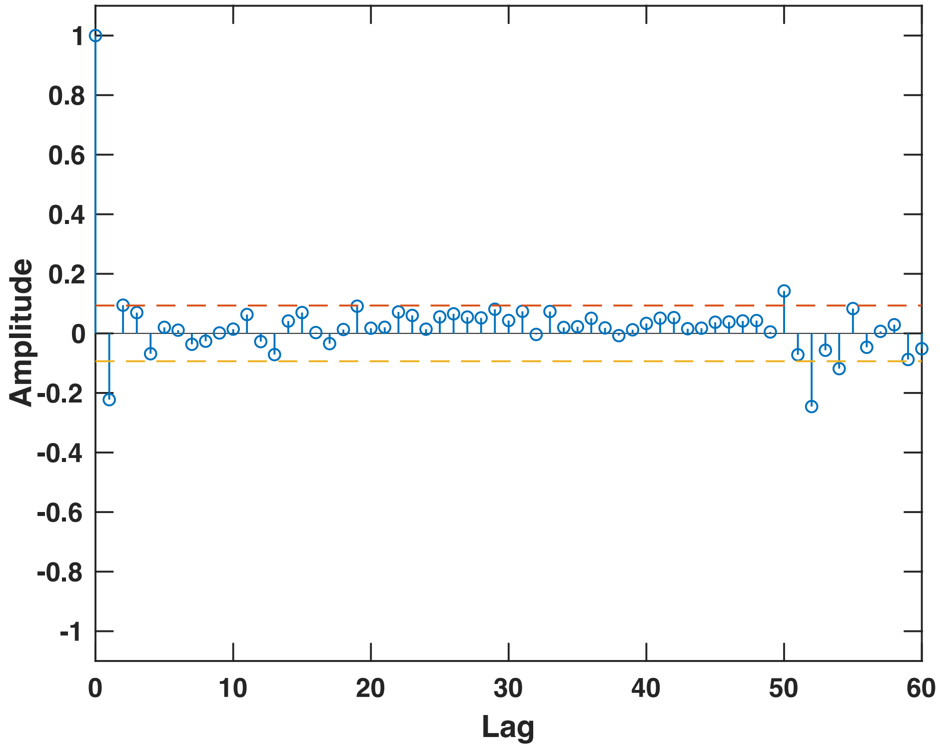
PACF of ∇y_t

Unsurprisingly, the data also has a strong annual cycle.

Mortality data



ACF of $(1 + a_{52}z^{-52})\nabla y_t$

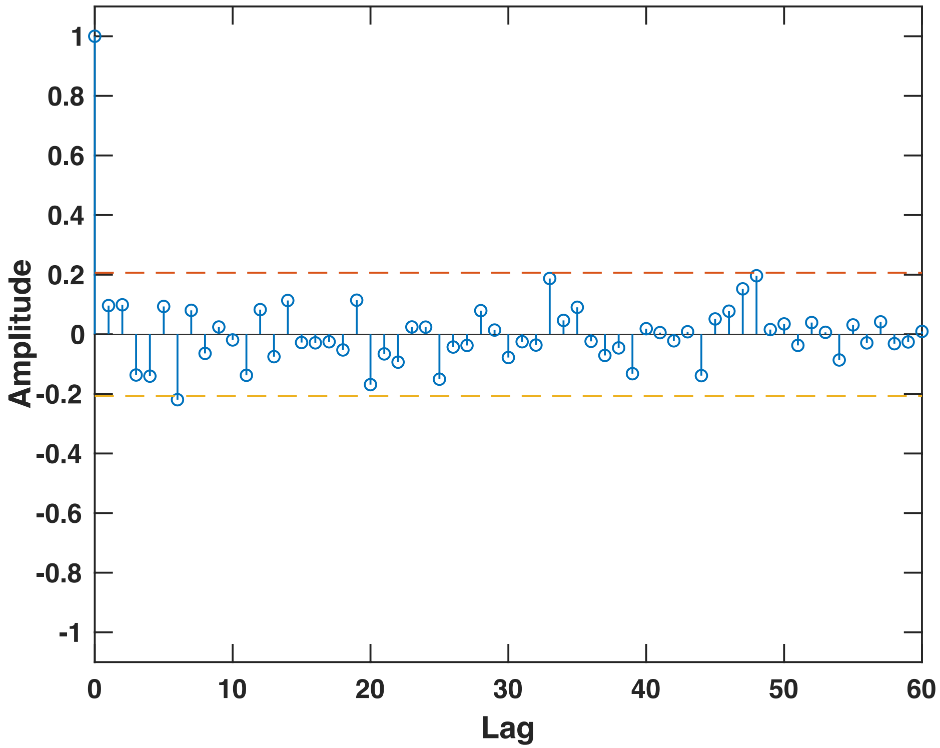


PACF of $(1 + a_{52}z^{-52})\nabla y_t$

The ACF suggests that we need to add a c_1 term.

Note: when modelling, we first model AR-components, than add MA-components, then return to the AR.

Mortality data



The residual of $(1 + a_{52}z^{-52})\nabla y_t = (1 + c_1 z^{-1})e_t$ is white, suggesting that this is a reasonable model.

Important: we need to re-estimate the a_{52} coefficient when adding the c_1 term. More on this soon.