The Box-Jenkins Methodology for ARIMA Model Selection

Identification step

- Examine the time plot of the series.
    - Identify outliers, missing values, and structural breaks in the data.
    - Non-stationary variables may have a pronounced trend or have changing variance.
    - Transform the data if needed. Use logs, differencing, or detrending.
        - Using logs works if the variability of data increases over time.
        - Differencing the data can remove trends. But over-differencing may introduce dependence when none exists.
- Examine the autocorrelation function (ACF) and partial autocorrelation function (PACF).
    - Compare the sample ACF and PACF to those of various theoretical ARMA models. Use properties of ACF and PACF as a guide to estimate plausible models and select appropriate p, d, and q.
    - With empirical data, several models may need to be estimated.
    - Differencing may be needed if there is a slow decay in the ACF.

Estimation step

- Estimate ARMA models and examine the various coefficients.
- The goal is to select a stationary and parsimonious model that has significant coefficients and a good fit.

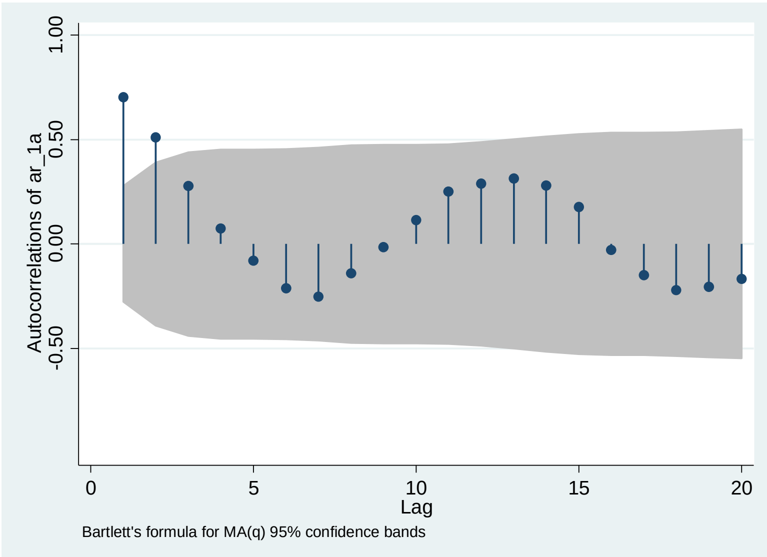
Diagnostic checking step

- If the model fits well, then the residuals from the model should resemble a while noise process.
  - Check for normality looking at a histogram of the residuals or by using a quantile-quantile (Q-Q) plot.
  - Check for independence by examining the ACF and PACF of the residuals, which should look like a white noise.
  - The Ljung-Box-Pierce statistic performs a test of the magnitude of the autocorrelations of the correlations as a group.
  - Examine goodness of fit using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Use most parsimonious model with lowest AIC and/or BIC.

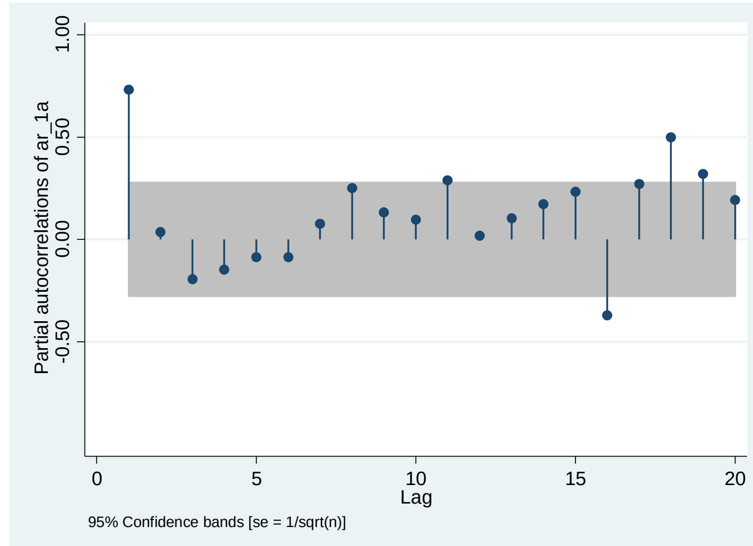
ACF and PACF properties

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

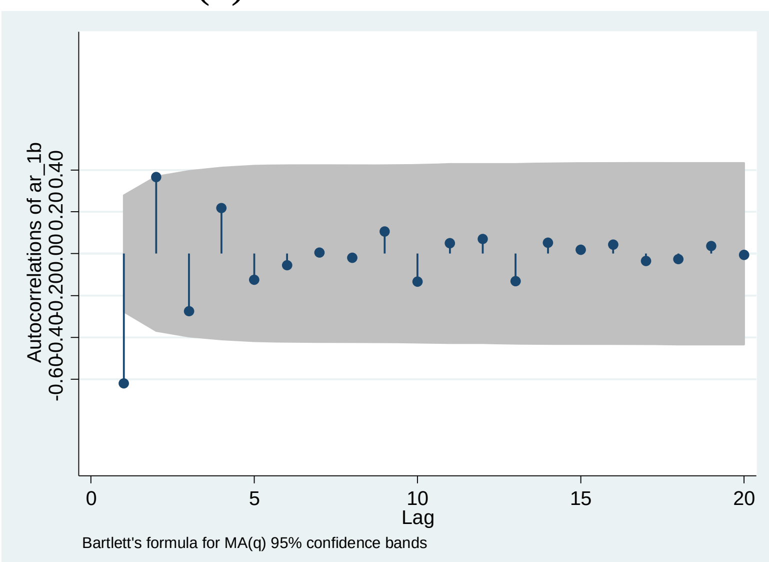
ACF of AR(1) with coefficient 0.8



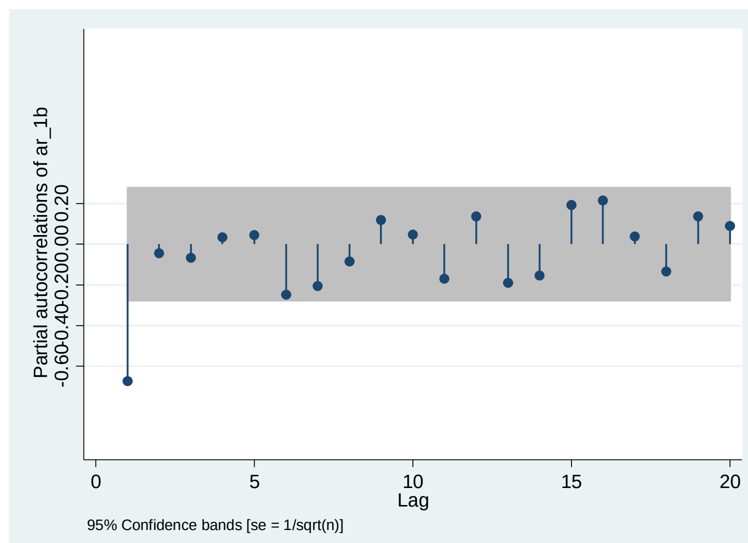
PACF of AR(1) with coefficient of 0.8



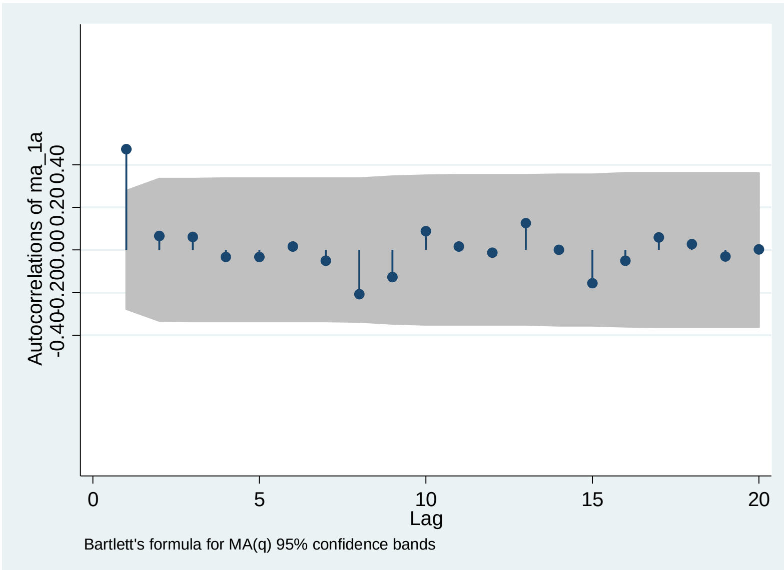
ACF of AR(1) with coefficient -0.8



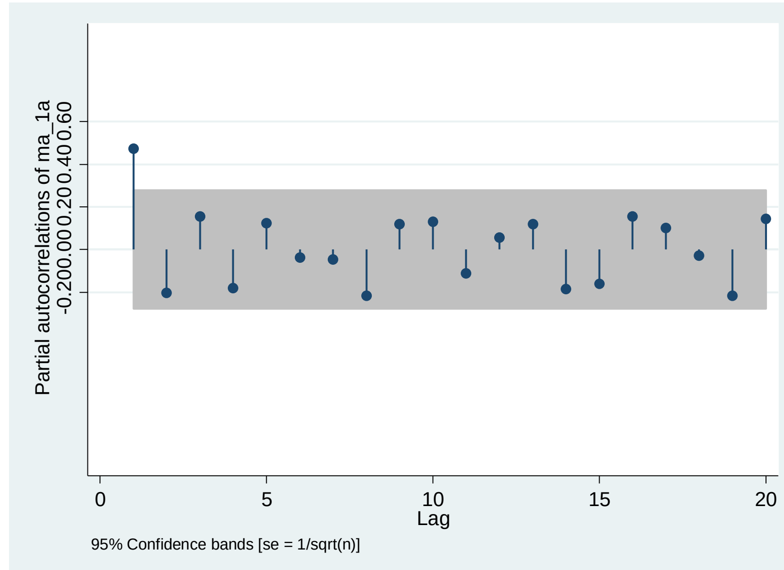
PACF of AR(1) with coefficient of -0.8



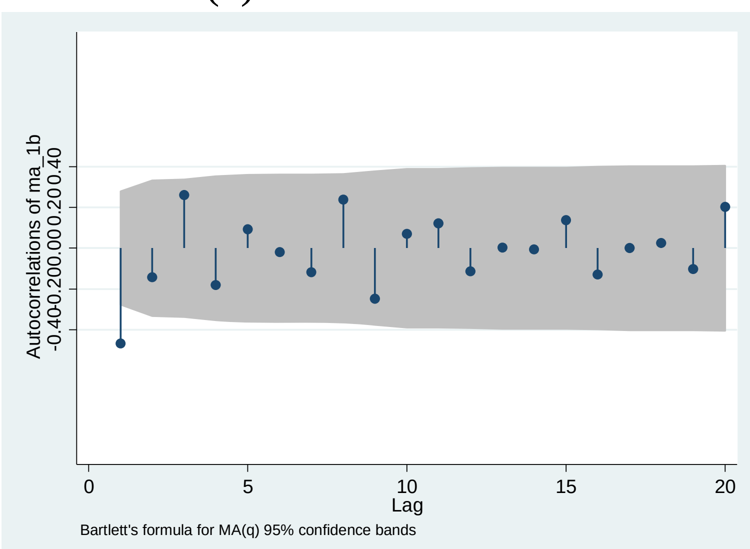
ACF of MA(1) with coefficient of 0.7



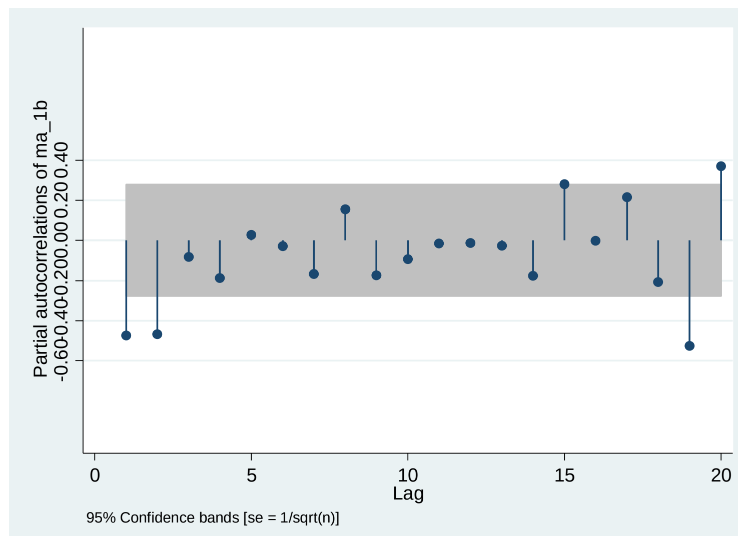
PACF of MA(1) with coefficient of 0.7



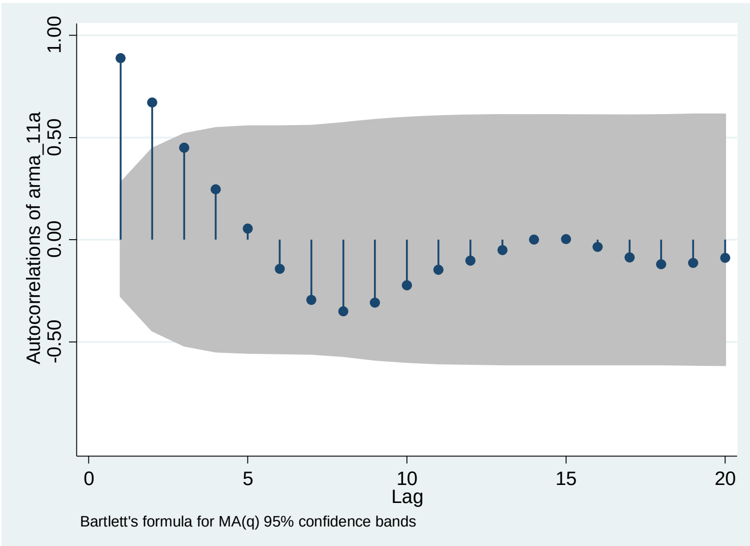
ACF of MA(1) with coefficient of -0.7



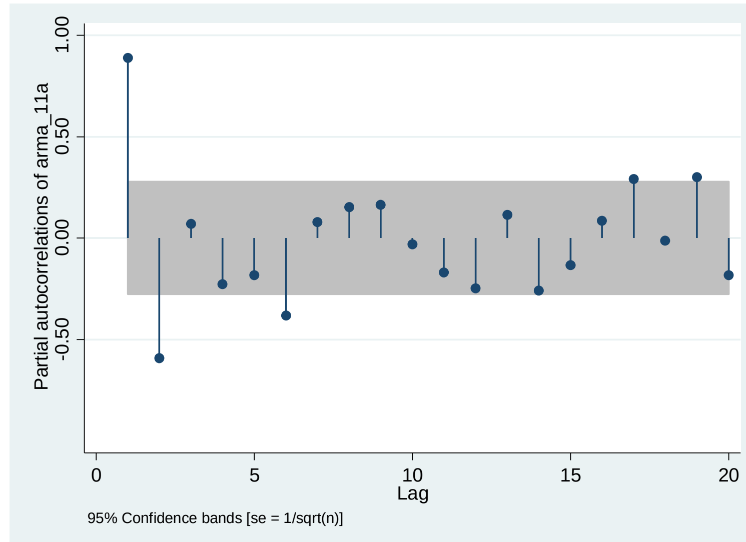
PACF of MA(1) with coefficient of -0.7



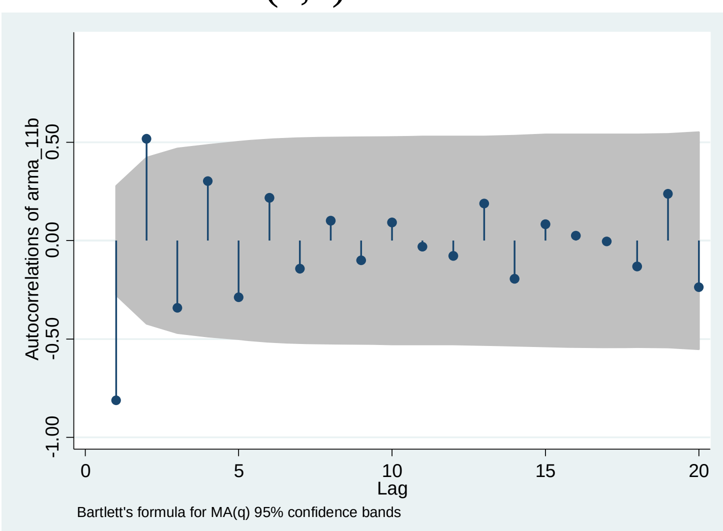
ACF of ARMA(1,1) with coeff 0.8 and 0.7



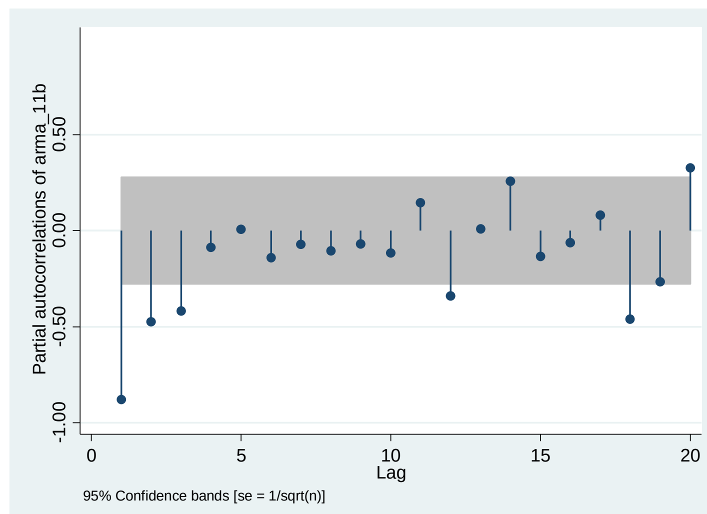
PACF of ARMA(1,1) with coeff 0.8 and 0.7



ACF of ARMA(1,1) with coeff −0.8 and −0.7



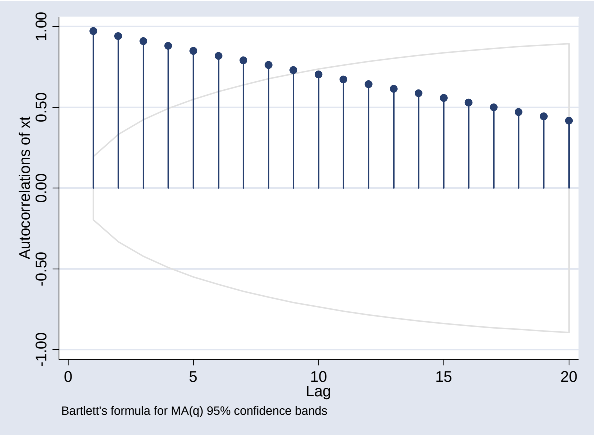
PACF of ARMA(1,1) with coeff −0.8 and −0.7



Seasonality

- Seasonality is a particular type of autocorrelation pattern where patterns occur every "season," like monthly, quarterly, etc.
- For example, quarterly data may have the same pattern in the same quarter from one year to the next.
- Seasonality must also be corrected before a time series model can be fitted.

ACF of non-stationary series - The ACF shows a slow decaying positive ACF.



ACF with seasonal lag (4) – ACF shows spikes every 4 lags.