Linear Time Series Models for Stationary data

Marius Ooms

Tinbergen Institute Amsterdam
Program

- **ARMA processes and properties**
  - Autoregressive processes, AR, PACF
  - Moving average processes, MA, invertibility
  - Mixed Autoregressive moving average processes, ARMA

- **Prediction and Forecasting**
  - AR
  - MA

- **Practice**
  - Identification ARMA orders
  - Estimation ARMA models
  - Test ARMA models
  - Prediction with ARMA model (see §7.1)
The **AR(1) process** is given by

\[ y_t = \phi y_{t-1} + \varepsilon_t \]  

leading to the particular solution (apply successive substitution: replace \( y_{t-1} \) by model for \( y_{t-1} \) using the constancy of the equation)

\[ y_t = \sum_{j=0}^{J-1} \phi^j \varepsilon_{t-j} + \phi^J y_{t-J}, \quad \text{with } J \text{ large.} \]

The **mean** is (treating \( y_{t-J} \) as a fixed number)

\[ \mathbb{E}(y_t) = \phi^J y_{t-J}, \]

s.t. we require \(|\phi| < 1\) for (asymptotic) stationarity.
AR(1) process, unconditional variance

Letting $J \to \infty$ leads to an (asymptotically) zero mean process and

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} = (1 - \phi L)^{-1} \varepsilon_t,$$

s.t. we have (unconditional) variance of an AR(1) process:

$$\gamma_0 = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{(1 - \phi^2)}.$$
AR(1) process, autocovariances

Assuming stationarity, there are two ways to obtain $\gamma_k$ for the AR(1) process (and higher order models)

1. via MA($\infty$) (book, theoretical),

2. via Yule-Walker equations, a practical time series analysis derivation, also applied in Method of Moments (MM) estimation and prediction)

1. Substitute the MA($\infty$) solution for $y_t$ to get:

$$
\gamma_k = E(y_t y_{t-k}) = E\left[ \left( \phi^k y_{t-k} + \sum_{j=0}^{k-1} \phi^j \varepsilon_{t-j} \right) y_{t-k} \right],
$$

leading to

$$
\gamma_k = \phi^k \gamma_0, \quad k \geq 0.
$$
AR process, autocovariance via Y-W

2. Alternatively derive the autocovariance using Yule Walker (Y-W) equations. Steps:
   1. Postmultiply both sides of (1) by $y_{t-k}$,
   2. take expectations and
   3. solve the systems of (Y-W) equations:

   $$
   \gamma_k = E(y_t y_{t-k}) = \phi E(y_{t-1} y_{t-k}) + E(\varepsilon_t y_{t-k}), \quad k \geq 0, 
   $$

   leading to

   $$
   \begin{align*}
   \gamma_k &= \phi \gamma(|k|-1), \quad |k| \geq 1 \\
   \gamma_0 &= \sigma^2_{\varepsilon}/(1 - \phi^2).
   \end{align*}
   $$

   So that the autocorrelation is $\rho_k = \phi^k$.

   *Exercise (1):* Show this expression also holds for negative $k$. 
AR(2) process

\[ \phi(L) y_t = \varepsilon_t, \quad \phi(L) = 1 - \phi_1 L - \phi_2 L^2 \]  \hspace{1cm} (2)

**Exercise (2):** Using Y-W equations and assuming stationarity, show for \( y_t \) in (2) that

\[ \rho_1 = \frac{\phi_1}{(1 - \phi_2)}, \]
\[ \rho_2 = \frac{\phi_1^2}{(1 - \phi_2)} + \phi_2, \]
\[ \rho_3 = \frac{\phi_1(\phi_1^2 + \phi_2)}{(1 - \phi_2)} + \phi_1 \phi_2, \]
\[ \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 3 : \text{linear difference equation!} \]

How do you derive \( \gamma_0 \)? Hint: use "Y-W" equation for \( k = 0 \).
Autocorrelation function (ACF) of an AR\((p)\) process is a solution to a **linear difference equation of order** \(p\). Refer to your mathematics book (e.g. Binmore and Davis (2001)), for the solution of linear difference equations with possibly complex and/or multiple roots. Linear difference equations are **not** covered in the appendices of Heij et al. (2004).

In general: The ACF of stationary AR\((p)\) process dies out **exponentially, oscillating** in case of negative or complex roots of \(\phi(z) = 0\).

The AR order, \(p\), is not easily derived from the ACF. To identify \(p\), we need a transformation of the ACF, namely the **PACF**.
**Partial Autocorrelation Function (PACF)**

**Partial Autocorrelation function (PACF)**

**Definition PACF:**
Steps for \( k = 1, 2, 3, \ldots \):

1. Consider the \( k \) Yule-Walker equations an AR(\( k \)) process.
2. Solve the \( k \) Yule-Walker equations, 1, \ldots, \( k \) for the \( k \) AR parameters, \( \phi_{k1}, \ldots, \phi_{kk} \), given autocorrelations up to order \( k \).
3. The **partial autocorrelation coefficient** of lag \( k \) is given by the last coefficient of the solution: \( \phi_{kk} \)

Example: \( \phi_{11} = \gamma_1 / \gamma_0 = \rho_1 \).

\( \phi_{kk} \) is interpreted as the **final least squares coefficient in a \( k \)-th order autoregression** applied to an infinitely long 'population' time series, see book: (7.12). NB: this is **not** the way to compute the PACF!

**Exercise(3):** Show that \( \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \)
### SACF/PACF Log Dow Jones

Example SACF and SPACF for Log Dow Jones Index Industrials
Weekly data, 1896-2004

**Correlogram of LOG(DJIND)**

Date: 02/05/04   Time: 16:08  
Included observations: 5596

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0.999 0.999</td>
<td></td>
<td>5590.1</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 0.998 -0.006</td>
<td></td>
<td>11172</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 0.998 -0.009</td>
<td></td>
<td>16746</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 0.997 -0.004</td>
<td></td>
<td>22312</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 0.996 -0.004</td>
<td></td>
<td>27870</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 0.995 -0.004</td>
<td></td>
<td>33420</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 0.994 -0.001</td>
<td></td>
<td>38961</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 0.993 -0.005</td>
<td></td>
<td>44494</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 0.993 -0.005</td>
<td></td>
<td>50018</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 0.992 -0.004</td>
<td></td>
<td>55534</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 0.991 0.002</td>
<td></td>
<td>61042</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 0.990 -0.002</td>
<td></td>
<td>66541</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 0.989 -0.001</td>
<td></td>
<td>72031</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 0.988 0.001</td>
<td></td>
<td>77514</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 0.988 0.009</td>
<td></td>
<td>82988</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 0.987 0.011</td>
<td></td>
<td>88453</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 0.986 0.002</td>
<td></td>
<td>93911</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 0.985 0.002</td>
<td></td>
<td>99361</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 0.984 -0.002</td>
<td></td>
<td>104803</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 0.983 0.002</td>
<td></td>
<td>110237</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## SACF/PACF dLog Dow Jones

**Example SACF and SPACF for Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004**

Correlogram of DLOG(DJIND)

<table>
<thead>
<tr>
<th>Date: 02/05/04   Time: 16:09</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included observations: 5595</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.021</td>
<td>0.021</td>
<td>2.5084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.044</td>
<td>0.044</td>
<td>13.432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.022</td>
<td>0.020</td>
<td>16.163</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-0.009</td>
<td>-0.012</td>
<td>16.596</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-0.008</td>
<td>-0.009</td>
<td>16.938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.006</td>
<td>0.007</td>
<td>17.154</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>0.010</td>
<td>0.011</td>
<td>17.733</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-0.019</td>
<td>-0.020</td>
<td>19.785</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.027</td>
<td>0.026</td>
<td>23.818</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>0.002</td>
<td>0.002</td>
<td>23.842</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>-0.015</td>
<td>-0.016</td>
<td>25.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.004</td>
<td>0.003</td>
<td>25.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>-0.008</td>
<td>-0.007</td>
<td>25.591</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>-0.024</td>
<td>-0.023</td>
<td>28.872</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>-0.008</td>
<td>-0.007</td>
<td>29.239</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>-0.025</td>
<td>-0.023</td>
<td>32.651</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>0.044</td>
<td>0.048</td>
<td>43.461</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>-0.011</td>
<td>-0.011</td>
<td>44.080</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>-0.021</td>
<td>-0.025</td>
<td>46.634</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>0.019</td>
<td>0.020</td>
<td>48.709</td>
</tr>
</tbody>
</table>
Computing and using the PACF

The Sample PACF is derived from the SACF, e.g.:

\[ \hat{\phi}_{22} = \hat{\rho}_2 - \hat{\rho}_1^2 \frac{1}{1 - \hat{\rho}_1^2}. \]

An alternative estimator for \( \hat{\phi}_{\text{OLS,22}} \) is derived from an OLS regression of \( y_t \) on a constant, \( y_{t-1} \) and \( y_{t-2} \). In practice, \( \hat{\phi}_{22} \) and \( \hat{\phi}_{\text{OLS,22}} \) are close in large samples of weakly stationary processes.

Application of Sample PACF for Identification:
The order of pure AR\( (p) \) processes is easily derived from the PACF, as

\[ \phi_{kk} = 0, \ k > p \]

In empirical work one has to take the sampling variability of the SPACF into account.
Moving average processes and invertibility

The finite order MA\((q)\) process is given by

\[
y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}, \quad \text{or} \quad y_t = \theta(L)\varepsilon_t.
\]

Finite order MA process is always stationary with variance

\[
\gamma_0 = \mathbb{E}(y_t^2) = \sigma^2 \varepsilon \left(1 + \theta^2_1 + \ldots + \theta^2_q\right).
\]

The autocovariance function is obtained as (cf. Wold representation):

\[
\gamma_k = \sigma^2 \varepsilon \left(\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q\right), \quad k = 1, \ldots, q.
\]

So that,

\[
\gamma_k = 0 \quad \text{for} \quad k = q + 1, q + 2, \ldots.
\]
Autocorrelation MA(q) and invertibility

The autocorrelation function is obtained by \( \rho_k = \frac{\gamma_k}{\gamma_0} \).

The ACF of an MA(\( q \)) process “cuts off” after lag \( q \).

**Exercise (4):** Show \( \rho_1 = \frac{\theta}{(1+\theta^2)} \) for an MA(1). Draw the solution with \( \rho_1 \) as a function of \( \theta \), \(-2 < \theta < 2\).

For an MA(1) there are two values for \( \theta \neq -1, 0, 1 \) giving the same \( \rho_1 \), say \( \theta = x \) and \( \theta = 1/x \).

What about \( \theta = 1 \)?

There are always multiple solutions for \( \theta(s) \) that correspond to the same (set of) \( \rho(s) \), This applies to all MA(\( q \)) process.

For the MA(1), we choose only the solution with \( |\theta| \leq 1 \). The corresponding invertible (causal) MA-model allows reconstruction of the past innovations from an infinitely long series for \( y_t \):

\[
\varepsilon_t = (1 + \theta L)^{-1} y_t.
\]
Mixed Autoregressive Moving Average Processes

The general ARMA\((p, q)\) process is \(\phi(L)y_t = \theta(L)\varepsilon_t\) where

\[
\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p
\]
\[
\theta(L) = 1 + \theta_1 L + \theta_2 L^2 - \ldots + \theta_q L^q.
\]

**Stationarity** requires the **AR roots** of \(\phi(z) = 0\) to lie outside the unit circle, and **invertibility** requires the same for the **MA roots** of \(\theta(z) = 0\).

Given these conditions, the **mixed** ARMA\((p, q)\) process may alternatively be expressed as a **pure** AR\((\infty)\) process or as a **pure** MA\((\infty)\) process (“Wold representation”) of infinite order, namely

\[
\pi(L)y_t = \theta^{-1}(L)\phi(L)y_t = \varepsilon_t \quad \text{or} \quad y_t = \phi^{-1}(L)\theta(L)\varepsilon_t = \psi(L)y_t
\]
ARMA(1,1) and use in macroeconomics

The lowest mixed process is the ARMA(1,1) is given by

\[ x_t = \phi x_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \]
\[ \varepsilon_t \sim NID(0, \sigma^2_{\varepsilon}), \quad t = 1, \ldots, n. \]

Schwert (1987, Journal of Monetary Economics, 73-103) provides several reasons to consider and MA-components in (growth rates of) (US) macroeconomic time series, so it’s definitely worth going “beyond AR”:

1. Measurement error,
2. Permanent income hypothesis,
3. Time aggregation and seasonal adjustment,
4. Rational multi-period forecast errors (interest rates or exchange rates).
5. Good fit to macroecononomic data. Most important argument for ARMA().
Prediction and Forecasting

**Prediction** is computing future values of $y_t$ (point predictions) and confidence intervals, that is for $y_{n+1}, y_{n+2}, \ldots$, given $Y_n = y_1, \ldots, y_n$.

Purpose of prediction: obtain optimal **linear minimum mean squared error predictions**, assuming the autocovariance structure of the process $y_t$ is known. Predictions are denoted by $\hat{y}_{n+1|n}, \hat{y}_{n+2|n}, \ldots$.

**Forecasting** error $e_{n+j} = y_{n+j} - \hat{y}_{n+j|n}$ is due to

(i) ignorance of future errors $\varepsilon_{n+j}$

(ii) ignorance of pre-sample errors $\varepsilon_0, \varepsilon_{-1}, \ldots$

(iii) uncertainty in estimates ARMA parameters (and parameters deterministic part)

(iv) Model misspecification
Prediction errors due to future innovations

The last 3 types of error are usually ignored. Statistical prediction analysis is confined to (i), (ii) and sometimes (iii). Under Gaussianity (normality) conditional expectations correspond to optimal linear forecasts and forecast(error)s have normal distributions.

We start with the AR(1) case, \( y_t = \phi y_{t-1} + \varepsilon_t \):

\[
\hat{y}_{n+1|n} = E(y_{n+1}|Y_n) = E(\phi y_n + \varepsilon_{n+1}|Y_n) = \phi y_n,
\]

s.t.

\[
e_{n+1} = y_{n+1} - \phi y_n = \varepsilon_{n+1}
\]

with properties \( E(e_{n+1}) = 0 \) and \( \text{SPE}(1) = \text{Var}(e_{n+1}) = \sigma^2_{\varepsilon} \).
Variance Prediction Errors

Then,

\[ \hat{y}_{n+2|n} = E(y_{n+2}|Y_n) = E(\phi y_{n+1} + \varepsilon_{n+2}|Y_n) = \phi^2 y_n, \]

s.t. \( e_{n+2} = y_{n+2} - \phi^2 y_n = \phi \varepsilon_{n+1} + \varepsilon_{n+2} \) with properties

\[ E(e_{n+2}) = 0 \quad \text{and} \quad \text{SPE}(2) = \text{Var}(e_{n+2}) = \sigma_\varepsilon^2 (1 + \phi^2). \]

For \( h \)-step ahead prediction we have,

\[ y_{n+h} = \phi^h y_n + \phi^{h-1} \varepsilon_{n+1} + \ldots + \varepsilon_{n+h}, \]

s.t. \( \hat{y}_{n+h} = \phi^h y_n \) and \( e_{n+h} = \phi^{h-1} \varepsilon_{n+1} + \ldots + \varepsilon_{n+h}. \)
Interval prediction AR(1)

It follows that $E(e_{n+h}) = 0$ and (cf. MA($\infty$) representation AR(1):

$$y_t = \psi(L)e_t$$

$$\text{SPE}(h) = \text{Var}(e_{n+h}) = \sigma^2_\varepsilon(1 + \phi^2 + \ldots + \phi^{2(h-1)})$$

As $h \to \infty$, $\hat{y}_{n+h} \to 0$ and

$$\text{Var}(e_{n+h}) = \frac{\sigma^2_\varepsilon}{(1 - \phi^2)}.$$ 

and these correspond with unconditional mean and variance of $y_t$. In a similar way, predictions for the AR($p$) model can be derived with extra administration.
Point and interval prediction MA($q$)

Point and Interval Prediction with an MA($q$) model is trivial if one assumes all past $\varepsilon_t$ to be known. E.g. take an MA(1) model $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$ and assume $\varepsilon_0 = 0$, so that $\varepsilon_1, \ldots, \varepsilon_n$ are known, since $y_1 = \varepsilon_1, y_2 = \varepsilon_2 + \theta y_1, \ldots$.

The one-step ahead forecast for MA(1) is then:

$$\hat{y}_{n+1|n} = E(y_{n+1}|Y_n) = E(\varepsilon_{n+1} + \theta \varepsilon_n | Y_n) = \theta \varepsilon_n,$$

s.t. $e_{n+1} = y_{n+1} - \theta \varepsilon_n = \varepsilon_{n+1}$ with properties

$$E(e_{n+1}) = 0 \quad \text{and} \quad \text{SPE}(1) = \text{Var}(e_{n+1}) = \sigma^2_\varepsilon.$$

Exercise (5) 7.4 c (p. 713): Derive the “mean and variance prediction function” of an ARMA(1,1) model.
Identification of ARMA model orders, ACF

Method 1 for Identification: use ACF and PACF
This method helps to understand the data, but can only lead to an ‘educated guess’.

Take variance of SACF $r_k$ and SPACF $\phi_{kk}$ into account when trying to identify a model from SACF and SPACF. For $r_1 = \hat{\phi}_{11}$ use ‘rule of thumb’ variance of $\frac{1}{n}$ under $H_0$ that process is white noise. Use the following known properties:

1. The ACF of a pure MA($q$) process “cuts off” after lag $q$.
2. The PACF of a pure AR($p$) process “cuts off” after lag $p$.
3. ACF of pure AR($p$) and mixed ARMA($p,q$) ($p > 0$) processes die out exponentially (after lag $q$ for mixed ARMA), oscillating in case of negative or complex roots of $\phi(z) = 0$. The AR order, $p$, is not easily derived from the ACF.
Identification of ARMA model orders, AIC

**Method 2: Minimize AIC or SIC**

This is a general statistical procedure, not confined to time series analysis.

1. Estimate a collection of models. Do not include models with $p$ and $q$ (too) large, ($p > 4$ and $q > 4$) unless you have a compelling reason (e.g. seasonal patterns).

2. Select model with the best trade-off between fit (residual sum of squared one-step-ahead forecasting errors) and number of parameters, according to AIC or SIC. Let $p^*$ be $p + q$, One minimizes $-2 \times$ the loglikelihood plus a penalty depending on $p^*$. 

   $$
   \text{AIC: } -2 \times l(p^*) + 2p^* \\
   \text{SIC: } -2 \times l(p^*) + p^* \log n
   $$

Note: in normal regression models:

$$
-2 \times l(p^*) = c + n \log(\sum e_i^2 / n), \text{ c.f. } §4.3.2 \text{ and } §5.2.1.
$$
Estimation ARMA models

1. OLS: What are regressors? see also §5.5.4.

2. Nonlinear least squares (NLS): uses $AR(\infty)$ parameterisation

$$\pi(L) = \theta^{-1}(L)\phi(L)$$

To estimate $\varepsilon_t$ Assumes fixed starting values for $y_t$ and $\varepsilon_t$.


4. Exact Gaussian Maximum Likelihood (EML). Takes likelihood for first observations into account and avoids assumptions about presample values for $\varepsilon_t$. (TSP, RATS, SAS, PcGive). Not (yet) in Eviews.
Practical dangers for ARMA estimation

- \( \phi(z) = 0 \) and \( \theta(z) = 0 \) should **not** be so flexible as to have common roots. How to avoid? Do not overspecify AR and MA parts simultaneously. Simplest example: fit ARMA(1,1) model to white noise \( (\phi = -\theta) \): inference on parameter estimates completely unreliable!

- NLS (and Eviews-) estimator and inference bad when \( \theta(1) \approx 0 \). The arbitrary assumptions about presample \( \varepsilon_t \) really hurt in this case. How to avoid? Do not “overdifference” the data. NLS not so problematic when \( \phi(1) = 0 \).

- Exact ML estimator and inference tricky when \( \phi(1) \approx 0 \). How to avoid? Do not ”underdifference” the data.
Testing and Evaluating ARMA models

• Check white noise assumption residuals, Ljung-Box test, LM test (Breusch Godfrey). If $H_0$ rejected: add AR or MA parameter, or add regressors §7.3.

• Check other assumptions using residuals (homoskedasticity, stationarity). If $H_0$ rejected allow for a changing mean §7.3, or variance §7.4, or both.

• Forecast performance out-of-sample. Assess empirical coverage of theoretical confidence intervals: count no. of observations outside confidence interval. Compare Root Mean Squared Prediction Error (RMSE), or Mean Absolute Prediction Error (MAE) with forecasts of benchmark models. If performance unsatisfactory: simplify model or allow for changing mean or variance in the model: model nonstationarity.