Econometrics II – Heij et al. Chapter 7.1

Linear Time Series Models for Stationary data

Marius Ooms

Tinbergen Institute Amsterdam

Program

- *Introduction Modelling philosophy Time Series Analysis*
- *Time series concepts and models for stationary time series*
  - *Stationarity, White Noise, Random Walk*
  - *Correlogram, ACF*
  - *Lags, differences and lag polynomials*
  - *Linear stochastic processes*

Introduction

Often, for (short run) **forecasting** purposes, a simple model that **describes** the behaviour of a **single variable** (or a small set of variables) **in terms of its own past values** is satisfactory.

Research shows that simple, linear univariate (ARIMA) time-series models based on just a few parameters often have a better forecasting performance than larger multivariate dynamic simultaneous-equations models (SEM), especially for **aggregated data** like **macroeconomic** series.

Intro: Relation DSEM/VAR <-> ARIMA

Zellner and Palm, (JoE (1974, pp 17-54)), provide a rationale for this: each single variable from a **joint** linear dynamic SEM, §7.7.2, or VAR, follows a **marginal** linear univariate ARIMA process, see §7.6.1. An ARIMA model does not accumulate misspecifications related to different equations and variables as in a multivariate dynamic SEM.

In section §7.1 - §7.3, we study tools of linear analysis of time-series data starting with key concepts in §7.1. **Model building philosophy** is now **specific to general** with cycles:

ARIMA Model building

1. Transform data to stationarity §7.3 (logs, differences, breaks)
2. Identify a parsimonious ARIMA model for stationary series, comparing model and data characteristics, **§7.1.3- §7.1.5**, §7.2.3
3. Estimate ARIMA model §7.2.2
4. Test ARMA model model, §7.2.4. If **not OK**, restart with 1. or 2.
5. Forecast with ARIMA model **§7.1.6**, §7.2.4 : conditional mean and variance, one-step and multi-step

Key concepts of linear time series analysis

Stationarity, second order / weak

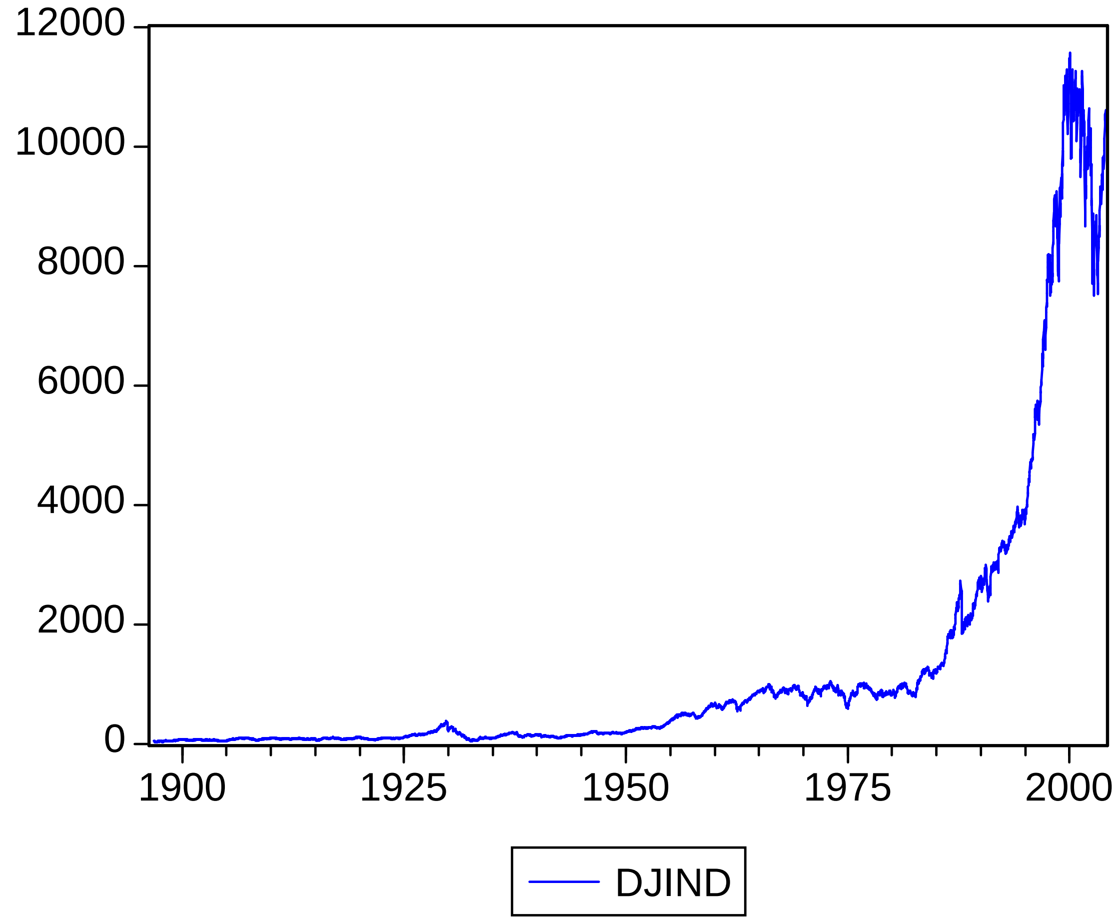
A stochastic process y_t is said to be stationary in mean and variance, when the following conditions are satisfied:

$$\mathrm{E}(y_t) = \mu, \quad \mathrm{Cov}(y_t, y_{t-k}) = \gamma_k,$$

for any t (that is, for any position in the time series) and for $k = 0, 1, 2, \ldots$.
This defines *second order (weak) stationarity*.

Dow Jones: Nonstationary series

Dow Jones Index Industrials Weekly 1896-2004



Key concept: SACF, White Noise (WN)

Consistent estimates for mean, variance and covariances for weakly stationary processes are obtained as follows:

$$\widehat{\mu} = \bar{y}, \quad \widehat{\gamma}_k = \frac{1}{n} \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}).$$

$\widehat{\gamma}_k$, $k = 0, 1, 2, \ldots$ is the **sample autocovariance function (SACF)**.

Simplest example of a **stationary** series is a **White Noise process (WN)** which we denote as ε_t. WN is a sequence of uncorrelated random variables with constant mean and variance:

$$\gamma_0 = \sigma_\varepsilon^2, \quad \gamma_k = 0, \quad k = 1, 2, \ldots.$$

Key concept: Random Walk (RW)

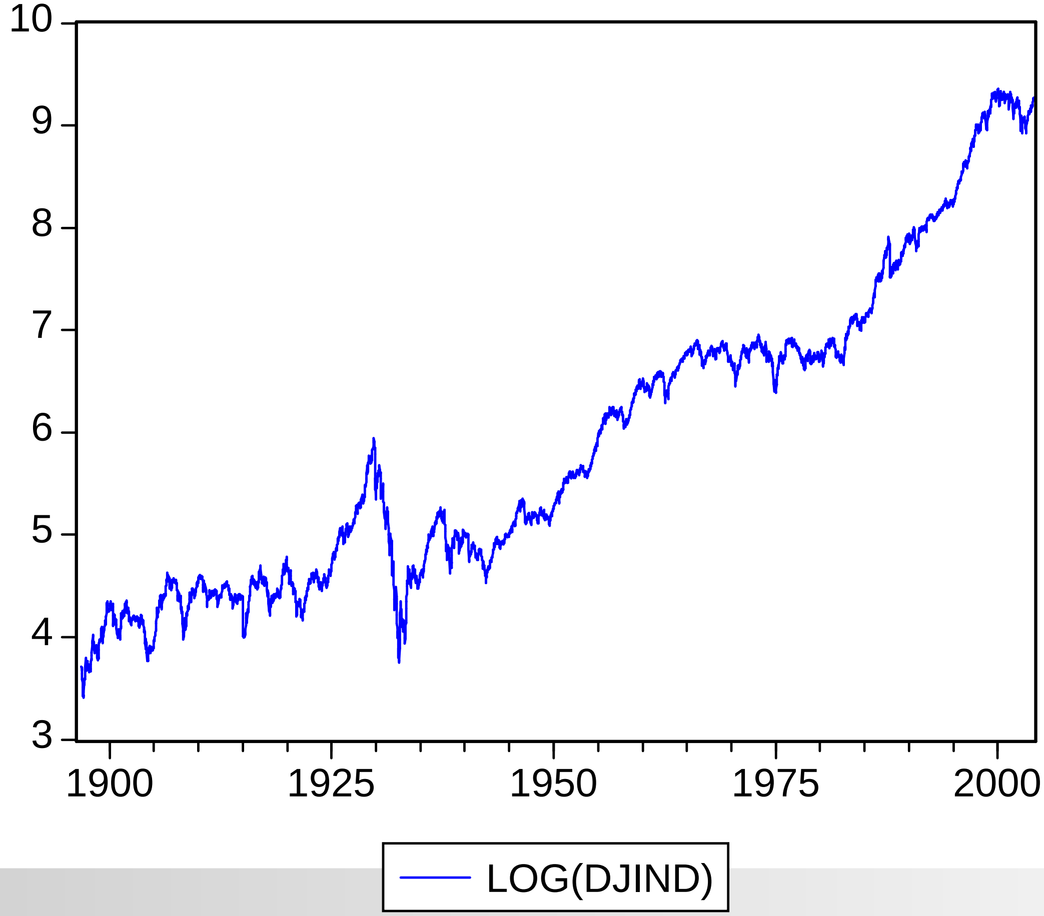
Best known economic example of a weakly **nonstationary** series is a **Random Walk Process (RW)**. RW is the partial sum process of a WN sequence:

$$u_t = \sum_{j=1}^{t} \varepsilon_j \quad t = 1, 2, \ldots .$$

The variance (conditional on starting values) of u_t increases (linearly) with t, even if $t \to \infty$, contradicting stationarity condition for variance.
Exercise (1) Derive this result.

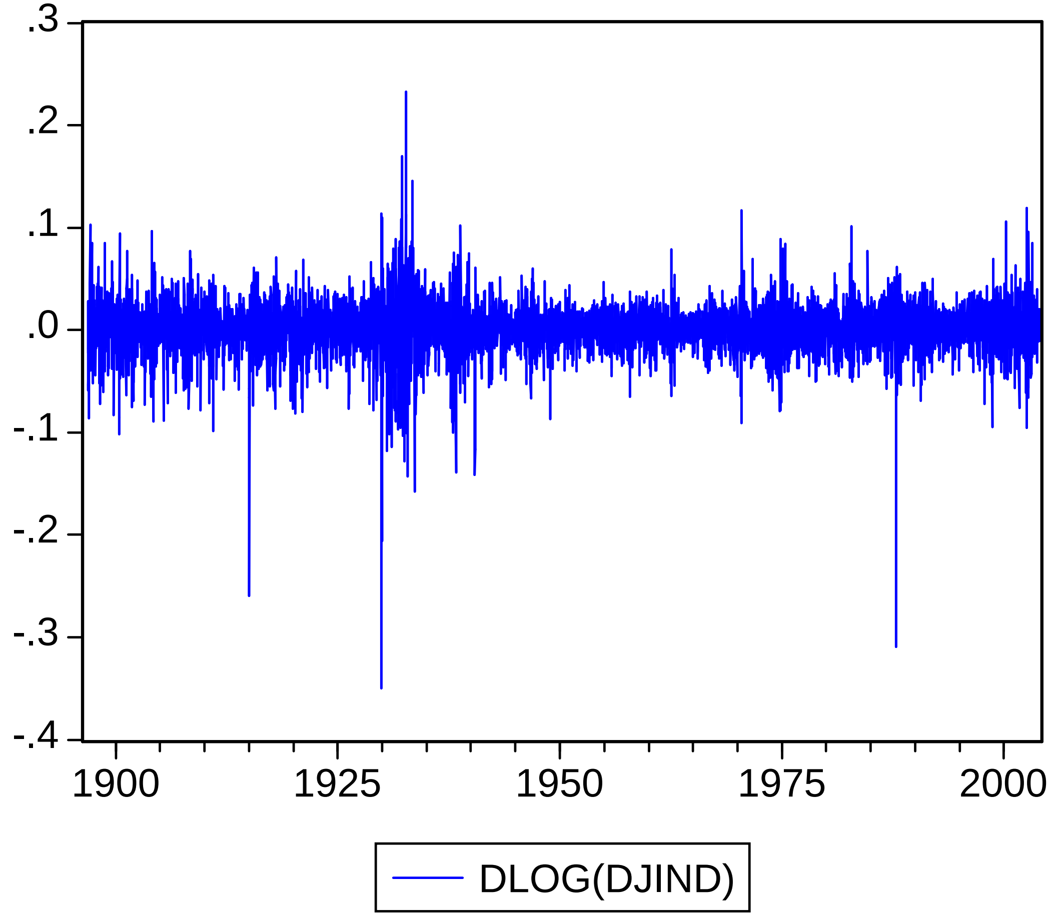
Log Dow Jones: Nonstationary series

Log transformed Dow Jones Index Industrials Weekly 1896-2004. *Exercise (2)*: approximate mean weekly growth.



Returns (DLog) DJI: Nonstationary series?

Returns Dow Jones Index Industrials Weekly 1896-2004



Autocovariance, -correlation, correlogram

The **autocovariance function**, γ_k, is seen as a function of lag k. One often plots the **autocorrelation function** which has the same shape and contains the same information on the dynamics, but is dimensionless:

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad k = 1, 2, \ldots .$$

Consistent estimates for weakly stationary processes are obtained by

$$\widehat{\rho}_k = r_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0},$$

a plot of these against $k = 1, 2, \ldots$ is known as the **correlogram**.

Identification / Model Selection principles

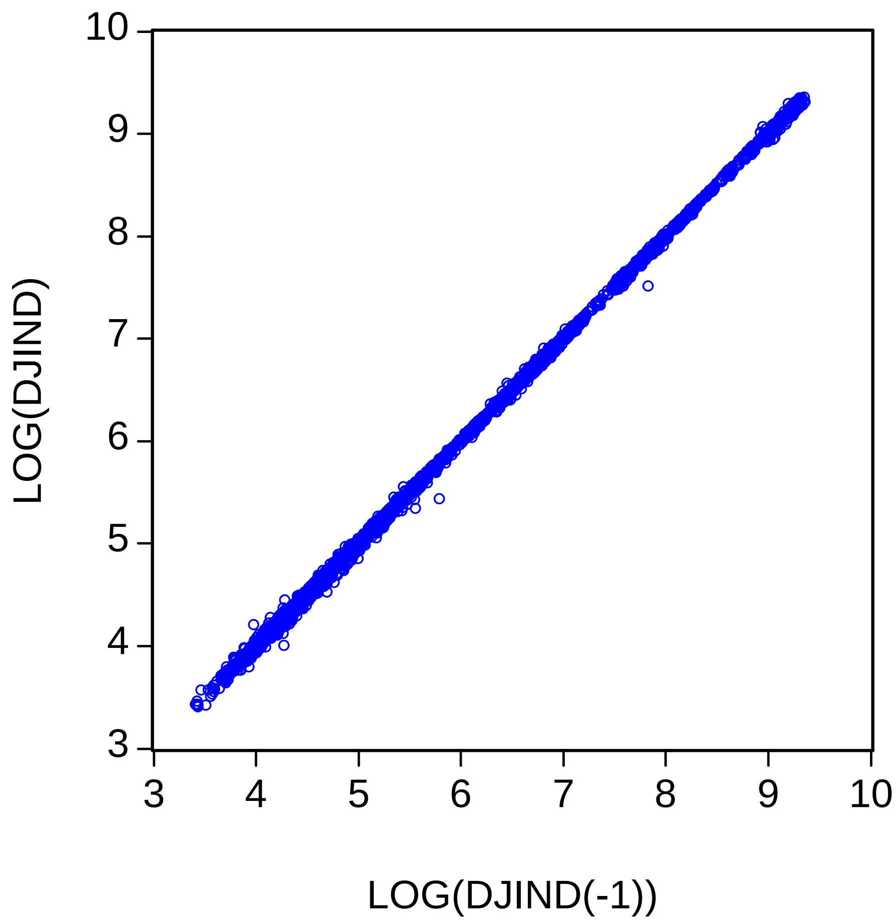
The sample characteristics of a time series should be similar to population characteristics of the statistical time series model. In the **identification** stage we choose our first model.

Time series plots and **correlograms** are the main tools for analysing the properties of time series. It is often also illuminating to consider (partial) **scatter** plots of y_t versus y_{t-k} in the context of the linear, infinite order autoregressive, (one-step-ahead) **prediction** decomposition of a time series process.

See also §5.5, where these techniques were applied as diagnostic tools on residuals.

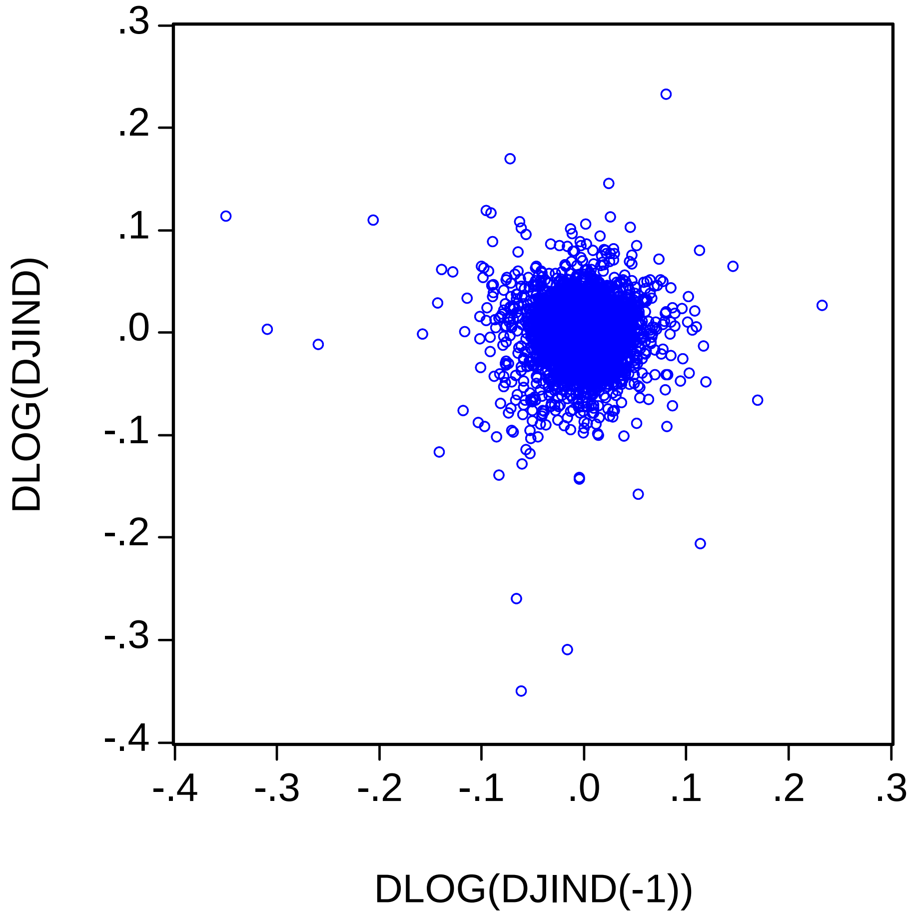
Scatter plot log Dow Jones vs. (-1)

Log Dow Jones Index Industrials Weekly 1896-2004



Scatter plot dlog Dow Jones vs (-1)

Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004



AR(∞) representation

Infinite order Autoregressive representation:

$$\begin{aligned} y_t &= E[y_t|Y_{t-1}] + \varepsilon_t \\ E[y_t|Y_{t-1}] &= \alpha + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots \qquad (7.2) \\ E[(y_t - \mu)|Y_{t-1}] &= \pi_1(y_{t-1} - \mu) + \pi_2(y_{t-2} - \mu) + \dots + \varepsilon_t \end{aligned}$$

where (here a constant) α s.t.
$(1 - \sum \pi_k)^{-1}\alpha = \mu$: **perfectly predictable** deterministic part;
$\pi_1(y_{t-1} - \mu) + (\pi_2 - \mu)y_{t-2} + \dots$: the (linearly) **predictable** stochastic part;
ε_t: **unpredictable** part of y_t, innovation process or prediction error, satisfying the White Noise condition.
Exercise (3): Using stationarity, show $E(y_t) = \mu = \frac{\alpha}{1-\sum \pi_k}$.

AR(20) Log Dow Jones

Log Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: LOGDJIND
Method: Least Squares
Date: 02/05/04 Time: 16:10
Sample(adjusted): 3/24/1897 1/27/2004
Included observations: 5576 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000576	0.001420	0.405789	0.6849
LOGDJIND(-1)	1.019829	0.013412	76.03649	0.0000
LOGDJIND(-2)	0.024193	0.019158	1.262802	0.2067
LOGDJIND(-3)	-0.028221	0.019144	-1.474136	0.1405
LOGDJIND(-4)	-0.027416	0.019114	-1.434339	0.1515
LOGDJIND(-5)	0.001765	0.019090	0.092460	0.9263
LOGDJIND(-6)	0.017536	0.019086	0.918799	0.3582
LOGDJIND(-7)	0.005282	0.019081	0.276802	0.7819
LOGDJIND(-8)	-0.034464	0.019080	-1.806301	0.0709
LOGDJIND(-9)	0.050894	0.019081	2.667203	0.0077
LOGDJIND(-10)	-0.024129	0.019091	-1.263928	0.2063
LOGDJIND(-11)	-0.021337	0.019091	-1.117659	0.2638
LOGDJIND(-12)	0.020287	0.019070	1.063788	0.2875
LOGDJIND(-13)	-0.007611	0.019065	-0.399203	0.6898
LOGDJIND(-14)	-0.019612	0.019065	-1.028677	0.3037
LOGDJIND(-15)	0.013889	0.019065	0.728512	0.4663
LOGDJIND(-16)	-0.013786	0.019066	-0.723046	0.4697
LOGDJIND(-17)	0.072059	0.019054	3.781789	0.0002
LOGDJIND(-18)	-0.059659	0.019080	-3.126773	0.0018
LOGDJIND(-19)	-0.015096	0.019095	-0.790558	0.4292
LOGDJIND(-20)	0.025666	0.013361	1.920898	0.0548

R-squared	0.999700	Mean dependent var	5.919391
Adjusted R-squared	0.999699	S.D. dependent var	1.496886
S.E. of regression	0.025960	Akaike info criterion	-4.460732
Sum squared resid	3.743740	Schwarz criterion	-4.435777
Log likelihood	12457.52	F-statistic	926489.2
Durbin-Watson stat	1.998905	Prob(F-statistic)	0.000000

AR(20) (+trend) Log Dow Jones

Log Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: LOGDJIND
Method: Least Squares
Date: 02/05/04 Time: 16:32
Sample(adjusted): 3/24/1897 1/27/2004
Included observations: 5576 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005196	0.002740	1.896455	0.0580
T	1.41E-06	7.13E-07	1.971410	0.0487
LOGDJIND(-1)	1.019106	0.013414	75.97385	0.0000
LOGDJIND(-2)	0.024200	0.019153	1.263498	0.2065
LOGDJIND(-3)	-0.028143	0.019139	-1.470451	0.1415
LOGDJIND(-4)	-0.027432	0.019109	-1.435569	0.1512
LOGDJIND(-5)	0.001766	0.019085	0.092521	0.9263
LOGDJIND(-6)	0.017515	0.019081	0.917944	0.3587
LOGDJIND(-7)	0.005277	0.019076	0.276605	0.7821
LOGDJIND(-8)	-0.034449	0.019075	-1.805995	0.0710
LOGDJIND(-9)	0.050860	0.019076	2.666130	0.0077
LOGDJIND(-10)	-0.024108	0.019086	-1.263142	0.2066
LOGDJIND(-11)	-0.021308	0.019086	-1.116404	0.2643
LOGDJIND(-12)	0.020246	0.019065	1.061934	0.2883
LOGDJIND(-13)	-0.007609	0.019060	-0.399232	0.6897
LOGDJIND(-14)	-0.019612	0.019060	-1.028949	0.3035
LOGDJIND(-15)	0.013859	0.019060	0.727116	0.4672
LOGDJIND(-16)	-0.013814	0.019061	-0.724709	0.4687
LOGDJIND(-17)	0.072054	0.019049	3.782509	0.0002
LOGDJIND(-18)	-0.059585	0.019075	-3.123700	0.0018
LOGDJIND(-19)	-0.015090	0.019090	-0.790459	0.4293
LOGDJIND(-20)	0.024892	0.013364	1.862693	0.0626

R-squared	0.999701	Mean dependent var	5.919391
Adjusted R-squared	0.999699	S.D. dependent var	1.496886
S.E. of regression	0.025954	Akaike info criterion	-4.461073
Sum squared resid	3.741122	Schwarz criterion	-4.434930
Log likelihood	12459.47	F-statistic	882829.4
Durbin-Watson stat	1.998889	Prob(F-statistic)	0.000000

AR(20) dLog Dow Jones

Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: DLOGDJIND
Method: Least Squares
Date: 02/05/04 Time: 16:11
Sample(adjusted): 3/31/1897 1/27/2004
Included observations: 5575 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000964	0.000352	2.735265	0.0063
DLOGDJIND(-1)	0.020379	0.013416	1.519045	0.1288
DLOGDJIND(-2)	0.044236	0.013414	3.297873	0.0010
DLOGDJIND(-3)	0.014778	0.013425	1.100752	0.2711
DLOGDJIND(-4)	-0.011092	0.013408	-0.827272	0.4081
DLOGDJIND(-5)	-0.009720	0.013402	-0.725309	0.4683
DLOGDJIND(-6)	0.008131	0.013385	0.607450	0.5436
DLOGDJIND(-7)	0.013087	0.013379	0.978222	0.3280
DLOGDJIND(-8)	-0.021609	0.013379	-1.615144	0.1063
DLOGDJIND(-9)	0.029758	0.013382	2.223753	0.0262
DLOGDJIND(-10)	0.005215	0.013385	0.389629	0.6968
DLOGDJIND(-11)	-0.016651	0.013385	-1.243954	0.2136
DLOGDJIND(-12)	0.004730	0.013382	0.353478	0.7237
DLOGDJIND(-13)	-0.003570	0.013365	-0.267144	0.7894
DLOGDJIND(-14)	-0.023138	0.013362	-1.731686	0.0834
DLOGDJIND(-15)	-0.008880	0.013364	-0.664465	0.5064
DLOGDJIND(-16)	-0.022613	0.013359	-1.692672	0.0906
DLOGDJIND(-17)	0.048747	0.013360	3.648812	0.0003
DLOGDJIND(-18)	-0.011406	0.013374	-0.852851	0.3938
DLOGDJIND(-19)	-0.025999	0.013360	-1.946003	0.0517
DLOGDJIND(-20)	0.020418	0.013361	1.528155	0.1265

R-squared	0.008920	Mean dependent var	0.001019
Adjusted R-squared	0.005351	S.D. dependent var	0.026027
S.E. of regression	0.025957	Akaike info criterion	-4.460958
Sum squared resid	3.742220	Schwarz criterion	-4.435999
Log likelihood	12455.92	F-statistic	2.499356
Durbin-Watson stat	2.000775	Prob(F-statistic)	0.000233

Practical purpose of linear time series analysis

If y_t is stationary, a nearly standard regression interpretation of (7.2) applies.

We strive to **minimize the variance** and **remove serial correlation in the "error term"**, ε_t, using **parsimonious** linear models for the **deterministic** part and the **predictable stochastic** part of y_t.
Exercise (4): Show that the Durbin-Watson statistic is a consistent estimator of $2(1-\rho_1)$, c.f. §5.5.3. Which of the three regression models above is optimal in this respect using standard regression criteria?

Some time series algebra

Lags, differences and lag polynomials

The **lag operator** L is useful in time series algebra:

$$Ly_t = y_{t-1}, \quad L^k y_t = y_{t-k}, \quad L^0 = 1, \quad L^{-k} y_t = y_{t+k} \quad \text{(lead)}.$$

The **first difference operator** is:

$$\Delta = 1 - L, \quad \Delta y_t = y_t - y_{t-1}.$$

Second differences:

$$\Delta^2 = \Delta\Delta = 1 - 2L + L^2.$$

Lag polynomials

Examples of *lag polynomials* are

$$
\begin{aligned}
\phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p, \\
\theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q,
\end{aligned}
$$

where p and q are some fixed maximum lag values.

The set of lag polynomials $\phi(L)$ is isomorphic to the set of algebraic polynomials $\phi(z)$, with $z \in \mathbb{C},\ |z| \leq 1$.

The **inverse of a lag polynomial** $\phi(L)$, denoted as $\phi^{-1}(L)$, is defined so that the product (convolution) is the "neutral element" 1.

Inverse of lag polynomial

In mathematical notation:

$$\phi^{-1}(z) \cdot \phi(z) = 1 =$$
$$\ldots + 0 \cdot z^{-2} + 0 \cdot z^{-1} + 1 \cdot z^{0} + 0 \cdot z^{1} + 0 \cdot z^{2} + \ldots$$

Example: $\phi(L) = 1 - \phi L$

$$\begin{aligned} (1 - \phi L)^{-1} &= 1L^0 + \phi L + \phi^2 L^2 + \ldots \\ &= \sum_{j=0}^{\infty} \phi^j L^j \end{aligned}$$

A **rational lag polynomial** $\phi^{-1}(L)\theta(L)$ is defined analogously.

Exercise (5): Derive the first 4 terms of $(1 - \phi_1 L - \phi_2 L^2)^{-1}$.

Linear Stochastic processes, Wold representation

A general representation of any zero-mean weakly stationary process y_t is the "linearly indeterministic component" of the so-called Wold-decomposition (MA(∞)):

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

with ε_t WN, where $\psi_0, \psi_1, \psi_2, \ldots$ are parameters for which

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty, \quad \text{or} \quad \sum_{j=0}^{\infty} |\psi_j| < \infty,$$

Variance/covariance MA(∞)

so that we have a finite variance and finite covariances

$$\gamma_k = \mathrm{E}[y_t y_{t-k}] = \mathrm{E}[\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \sum_{h=0}^{\infty} \psi_h \varepsilon_{t-h-k}] = \sigma_\varepsilon^2 \sum_{h=0}^{\infty} \psi_{k+h} \psi_h$$

Neither this representation, nor the infinite order autoregression are useful in practice as it is impossible to estimate all the parameters from a finite sample.

The properties of a linear stochastic process are expressed in the autocovariance function. All stationary stochastic linear processes can be efficiently approximated by the ARMA model class, which uses a parsimonious rational approximation for $\psi(L)$ or $\pi(L)$.

Econometrics II – Heij et al. Chapter 7.1/7.2

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)*

Autoregressive process of order 1, mean

The **AR(1) process** is given by $y_t = \phi y_{t-1} + \varepsilon_t$ leading to the particular solution (successive substitution)

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

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

$$\mathrm{E}(y_t) = \phi^J y_{t-J},$$

s.t. we require $|\phi| < 1$ for (asymptotic) stationarity.

AR(1) process, 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 variance

$$\gamma_0 = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}.$$

AR(1) process, autocovariances

Assuming stationarity, there are two ways to obtain γ_k for the AR(1) process, 1. via MA(∞), 2. via Yule Walker
1. Substituting solution for y_t :

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

leading to

$$\gamma_k = \phi^k \gamma_0, \quad k \geq 0.$$

AR process, autocovariance via Y-W

2. Alternatively using Yule Walker (Y-W) equations:

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

leading to

$$\begin{aligned} \gamma_k &= \phi\gamma(|k|-1), \qquad |k| \geq 1 \\ \gamma_0 &= \sigma_\varepsilon^2/(1-\phi^2). \end{aligned}$$

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

Exercise(1): Show this expression holds for negative k.

AR(2) process

$$\phi(L)y_t = \varepsilon_t, \quad \phi(L) = 1 - \phi_1 L - \phi_2 L^2,$$

Exercise(2): Using Y-W equations, show for this AR(2) that

$$\begin{aligned}
\rho_1 &= \frac{\phi_1}{(1-\phi_2)}, \\
\rho_2 &= \frac{\phi_1^2}{(1-\phi_2)} + \phi_2, \\
\rho_3 &= \frac{\phi_1(\phi_1^2+\phi_2)}{(1-\phi_2)} + \phi_1\phi_2, \\
\rho_4 &= \phi_1\rho_3 + \phi_2\rho_2 \quad \text{etc.}
\end{aligned}$$

How do you derive γ_0? Hint: use "Y-W" equation for $k = 0$.

Form of ACF AR(p) process

Autocorrelation function (ACF) of and AR(p) process is a solution to a linear difference equation of order p.

In general: ACF of 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 ACF. To identify p, we need a transformation of the ACF, namely the PACF.

Partial Autocorrelation Function (PACF)

Partial Autocorrelation function

Definition:
Consider k Yule-Walker equations written up for an AR(k) process. 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.

The **partial autocorrelation coefficient** of lag k is given by the solution ϕ_{kk} (so $\phi_{11} =\gamma_1$/γ_0 = ρ_1)

ϕ_{kk} can be interpreted just as the **final least squares coefficient in a k-th order autoregression** applied to an infinitely long 'population' time series, see book: (7.12).

Exercise(3): Show that $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$

SACF/PACF Log Dow Jones

Log Dow Jones Index Industrials Weekly 1896-2004

Correlogram of LOG(DJIND)

Date: 02/05/04 Time: 16:08
Sample: 5/27/1896 1/27/2004
Included observations: 5596

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

SACF/PACF dLog Dow Jones

Returns (dlog) Dow Jones Index Industrials Weekly 1896-2004

Correlogram of DLOG(DJIND)

Date: 02/05/04 Time: 16:09
Sample: 5/27/1896 1/27/2004
Included observations: 5595

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.021	0.021	2.5084	0.113
		2	0.044	0.044	13.432	0.001
		3	0.022	0.020	16.163	0.001
		4	-0.009	-0.012	16.596	0.002
		5	-0.008	-0.009	16.938	0.005
		6	0.006	0.007	17.154	0.009
		7	0.010	0.011	17.733	0.013
		8	-0.019	-0.020	19.785	0.011
		9	0.027	0.026	23.818	0.005
		10	0.002	0.002	23.842	0.008
		11	-0.015	-0.016	25.109	0.009
		12	0.004	0.003	25.209	0.014
		13	-0.008	-0.007	25.591	0.019
		14	-0.024	-0.023	28.872	0.011
		15	-0.008	-0.007	29.239	0.015
		16	-0.025	-0.023	32.651	0.008
		17	0.044	0.048	43.461	0.000
		18	-0.011	-0.011	44.080	0.001
		19	-0.021	-0.025	46.634	0.000
		20	0.019	0.020	48.709	0.000

Estimating and using the PACF, use of PACF

The sample PACF can be derived from the SACF, e.g.:

$$\widehat{\phi}_{22} = \frac{\widehat{\rho}_2 - \widehat{\rho}_1^2}{1 - \widehat{\rho}_1^2}.$$

An alternative estimator for ϕ_{22} is derived from OLS of y_t on a constant, y_{t-1} and y_{t-2}. $\hat{\phi}_{22}$ and $\hat{\phi}_{\text{OLS},22}$ are close in moderately large samples if the data are weakly stationary.

Application of Sample PACF for Identification:
The PACF is used to determine the order of AR(p) processes, as $\phi_{kk} = 0, k > p$.
AR order p is therefore easily derived from PACF.

Moving average processes and invertibility

The 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, zero mean, and variance

$$\gamma_0 = \mathrm{E}\left(y_t^2\right) = \sigma_\varepsilon^2 \left(1 + \theta_1^2 + \ldots + \theta_q^2\right).$$

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

$$\gamma_k = \sigma_\varepsilon^2(\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q), \quad k = 1, \ldots, q.$$

So that, $\gamma_k = 0$ for $k = q+1, q+2, \ldots$.

Autocorrelation MA(q) and invertibility

The autocorrelation function is obtained by $\rho_k = \frac{\gamma_k}{\gamma_0}$.
The ACF of and MA(q) process "cuts off" after lag q.
Exercise (4): Show $\rho_1 = \frac{\theta}{(1+\theta^2)}$ for an MA(1).

For an MA(1) there are two values for $\theta \neq -1, 0, 1$ giving the same ρ_1, say $\theta = x$ and $\theta = 1/x$.
graph . What about $\theta = 1$?

The multiple solutions for $\theta(s)$ given $\rho(s)$ apply to any MA(q) process. For the MA(1), we choose only solution with $|\theta| \leq 1$. The corresponding invertible (causal) MA-model allows reconstruction of the *innovations* from an infinite 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 roots of $\phi(z) = 0$ to lie outside the unit circle, and invertibility requires the same for the roots of $\theta(z) = 0$. Given these conditions, the ARMA(p, q) process may alternatively be expressed as a pure AR(∞) process or as a pure MA(∞) process ("Wold representation") of infinite order, namely

$$\theta^{-1}(L)\phi(L)y_t = \varepsilon_t \quad \text{or} \quad y_t = \phi^{-1}(L)\theta(L)\varepsilon_t.$$

ARMA(1,1)

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 other reasons for presence of MA-components in (US) macroeconomic time series, so it's definitely worth going ”beyond AR”.
Next to Zellner-Palm argument: Measurement error, permanent income hypothesis, time aggregation, rational multi-period forecast errors (interest rates or exchange rates). Most important for ARIMA(): good fit to data.

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: obtain optimal **linear minimum mean squared error predictions**. Predictions are denoted by $\widehat{y}_{n+1|n}, \widehat{y}_{n+2|n}, \ldots$.

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

(i) ignorance of future errors ε_{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 conditional expectations correspond to optimal forecasts and forecast(error)s have normal distributions.

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

$$\widehat{y}_{n+1|n} = \mathrm{E}(y_{n+1}|Y_n) = \mathrm{E}(\phi y_n + \varepsilon_{n+1}|Y_n) = \phi y_n,$$

s.t.

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

with properties $\mathrm{E}(e_{n+1}) = 0$ and SPE$(1) = \mathrm{Var}(e_{n+1}) = \sigma_\varepsilon^2$.

Variance Prediction Errors

Then,

$$\widehat{y}_{n+2|n} = \mathrm{E}(y_{n+2}|Y_n) = \mathrm{E}(\phi y_{n+1} + \varepsilon_{n+2}|Y_n) = \phi^2 y_n,$$

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

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

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. $\widehat{y}_{n+h} = \phi^h y_n$ and $e_{n+h} = \phi^{h-1}\varepsilon_{n+1} + \ldots + \varepsilon_{n+h}$.

Interval prediction AR(1)

It follows that $\mathrm{E}(e_{n+h}) = 0$ and (cf. MA(∞) representation AR(1): $y_t = \psi(L)\varepsilon_t$)

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

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

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

and these correspond with unconditional mean and variance of y_t. In a similar way, predictions for 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 nearly trivial if all past ε_t are known. E.g. take 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.

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

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

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

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

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: 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 ϕ_{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

1. The ACF of MA(q) process "cuts off" after lag q.
2. The PACF of AR(p) process "cuts off" after lag p.
3. ACF of AR(p) and ARMA(p,q) ($p > 0$) processes die out exponentially after lag q, oscillating in case of negative or complex roots of $\phi(z) = 0$. The AR order, p, is not easily derived from ACF.

Identification of ARMA model orders, AIC

Method 2: Minimize AIC or SIC

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

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

Note: in normal (nonlinear) regression models: $-2 * l(p^*) = c + n \log(\sum e_t^2/n)$, c.f. §4.3.2 and §5.2.1.

Estimation ARMA models

1. OLS: What are regressors? see also §5.5.4.
2. Nonlinear least squares (NLS): uses AR(∞) parameterisation $\pi(L) = \theta^{-1}(L)\phi(L)$ to estimate ε_t Assumes fixed starting values for y_t and ε_t.
3. Conditional (Gaussian) Maximum Likelihood (CML), conditional on $p + q$ starting values. Equals OLS for AR models. Beware of interpretation regression parameters. Loose observations at beginning sample.
4. Exact Gaussian Maximum Likelihood (EML). Takes likelihood for first observations into account.(TSP, RATS, SAS, PcGive). Not (yet) in Eviews.

Dangers for ARMA estimation

Dangers for methods:

- $\phi(z) = 0$ and $\theta(z) = 0$ should **not** have common roots. Do not overspecify AR and MA part simultaneously.
- NLS (and Eviews-) estimator and inference bad when $\theta(1) \approx 0$. Do not “overdifference” the data. NLS not so problematic when $\phi(1) = 0$.
- EML estimator and inference tricky when $\phi(1) \approx 0$. 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.