Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	    • Autocorrelation Function
	    • Moving Average and Autoregressive Models
	    • Partial Autocorrelation Function
	    • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• **Introduction** • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Definition Stationarity:

The underlying stochastic process (mean, variance, and covariance) is assumed to be **invariant with respect to time**.

$$\mu_y = E(y_t) = E(y_{t+m})$$

$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right] = E\left[(y_{t+m} - \mu_y)^2\right]$$

$$\gamma_k = Cov(y_t, y_{t+k}) = E\left[(y_t - \mu_y)(y_{t+k} - \mu_y)\right] = Cov(y_{t+m}, y_{t+m+k})$$

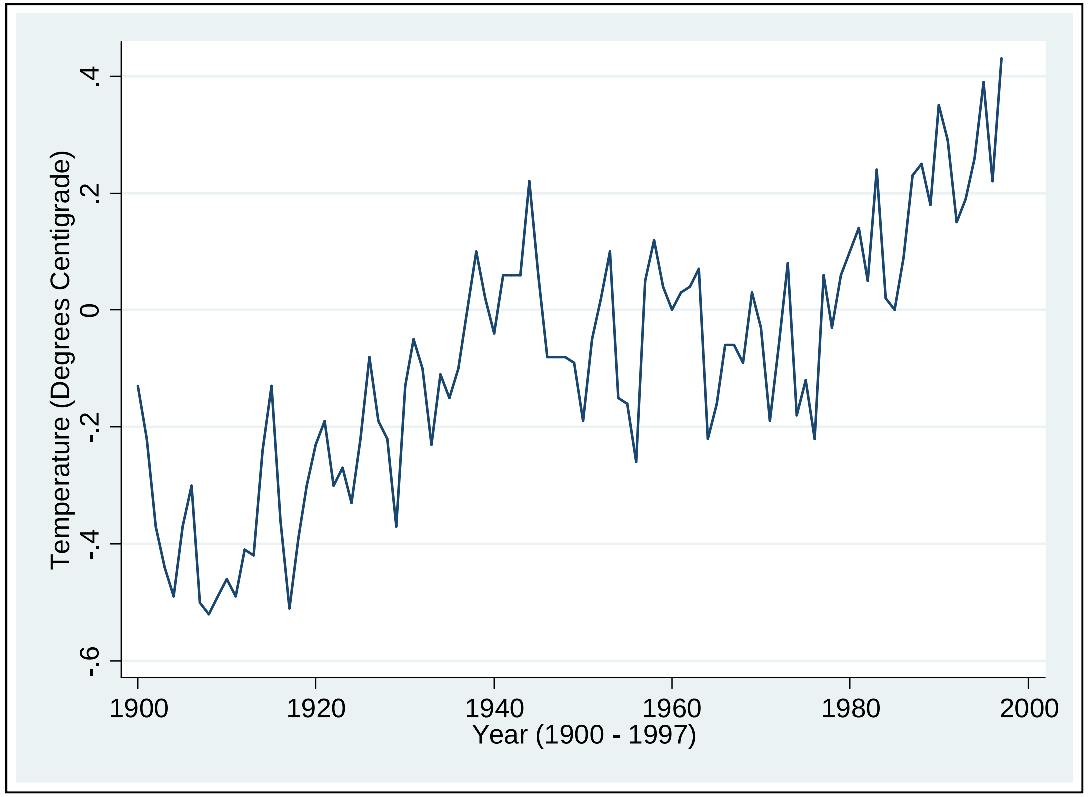
for any *t, k,* and *m*

Implications of Stationarity:

- no trend
- variance (magnitude of fluctuations) constant
- pattern of serial correlation does not change

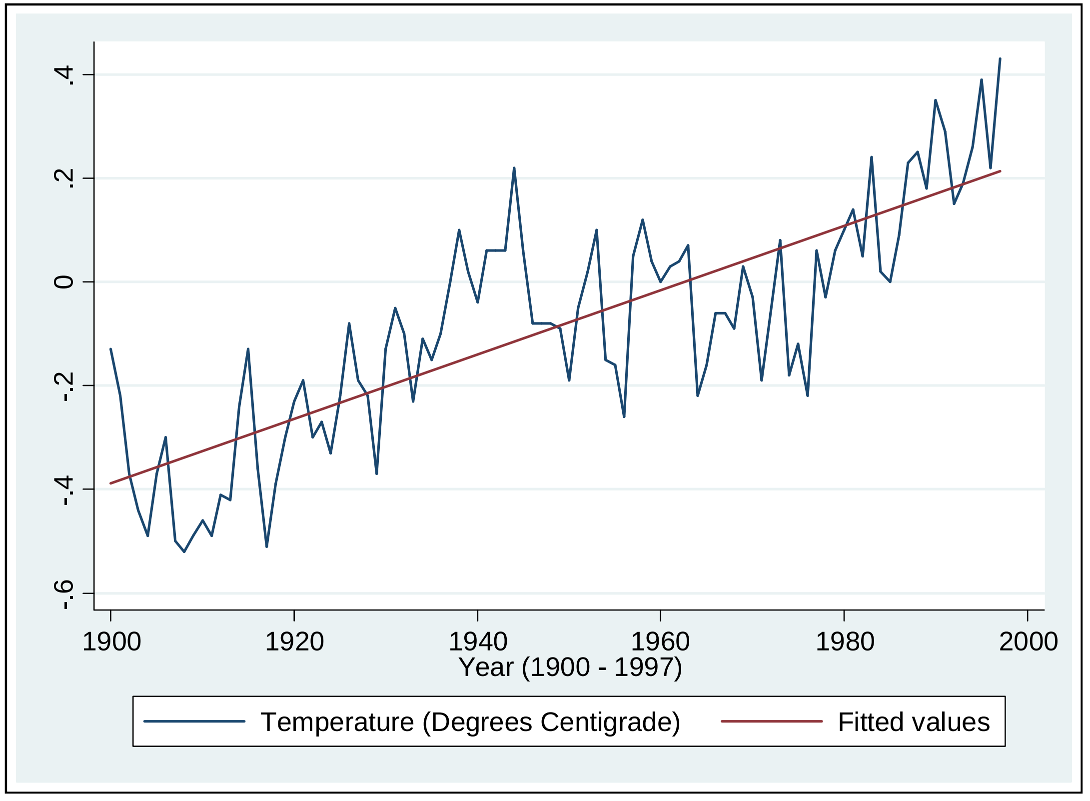
We could look at the series to try to detect deviations from stationarity.

Example: Global warming data



Shumway/Stoffer (2000) “Time Series Analysis and its Applications”

Example: Global warming data



Shumway/Stoffer (2000) “Time Series Analysis and its Applications”

Implications of Stationarity:

- no trend
- variance (magnitude of fluctuations) constant
- pattern of serial correlation does not change

Nonstationarity

- could arise from any deviation of the above
- most obvious and important deviation is a trend

We will focus on nonstationarity due to trending that can be ‘cured´ by eliminating the trend and apply ARMA models to the detrended series.

Nonstationarity: Trending

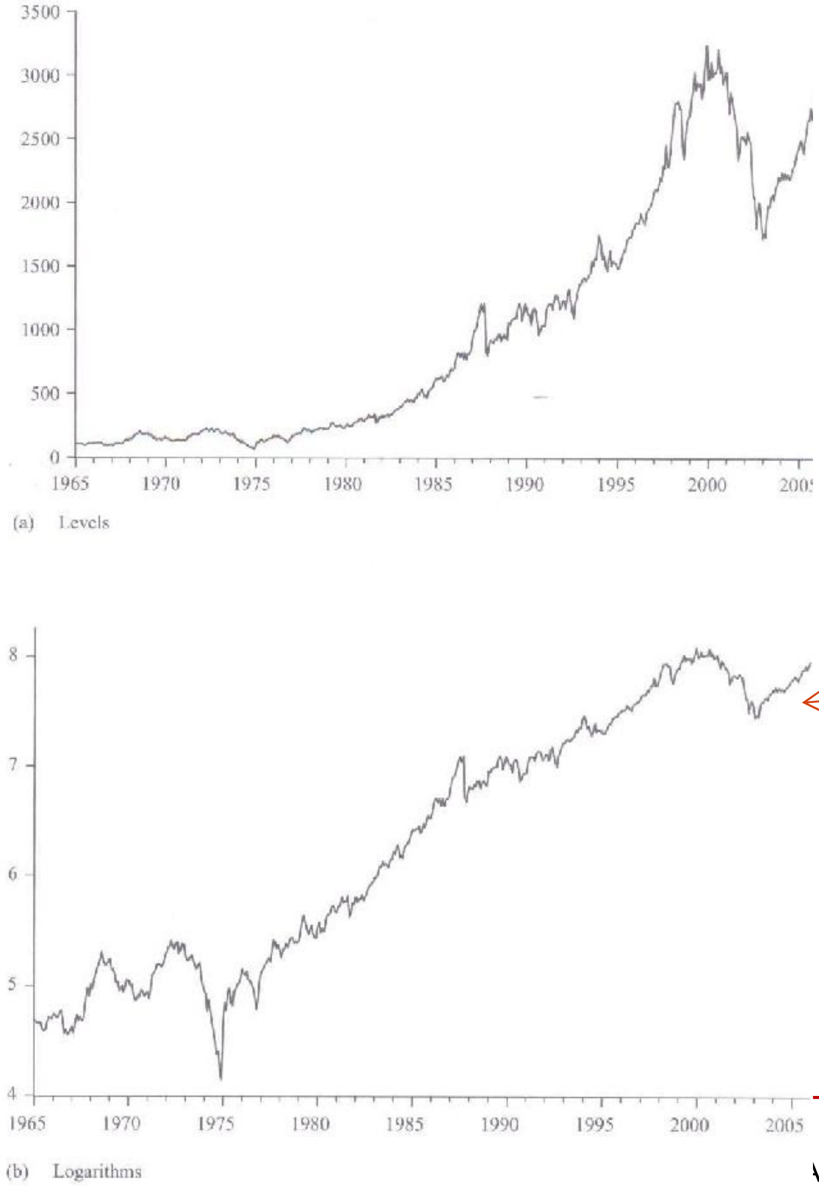


Figure 2.14 FTA All Share index (monthly 1965–2005)

Example:

monthly FTA All Share Index (1965-2005)

$$r_t = \log \frac{P_t}{P_{t-1}} = p_t - p_{t-1}$$

Nonstationarity: Trending ?

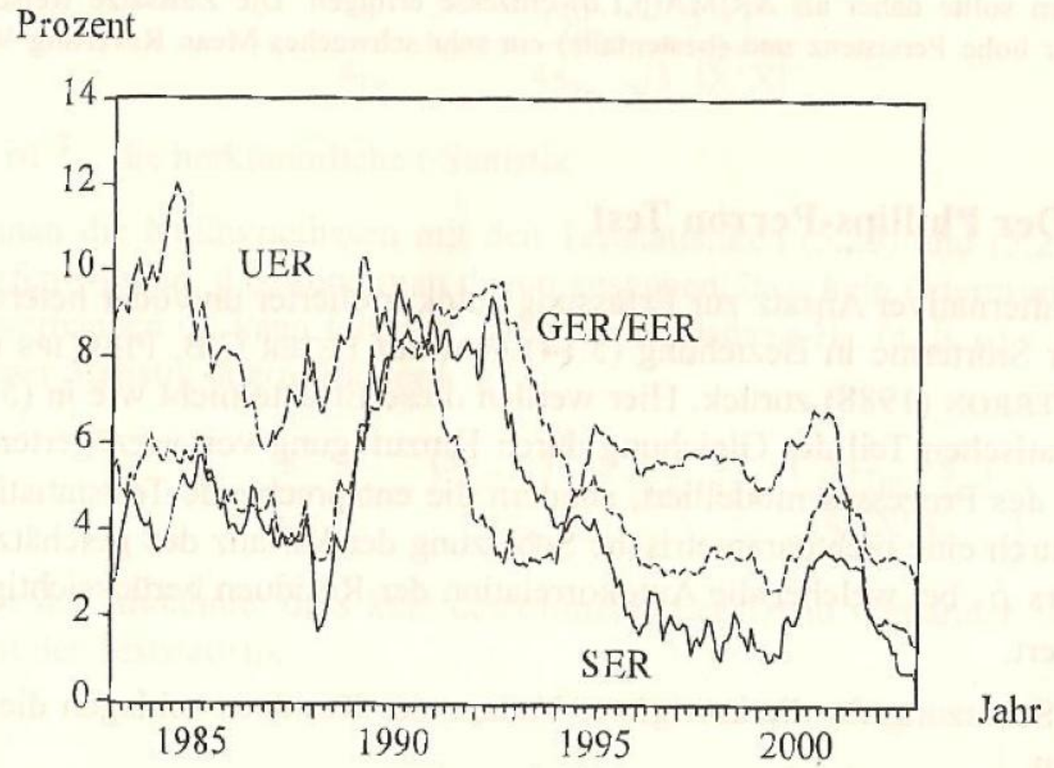
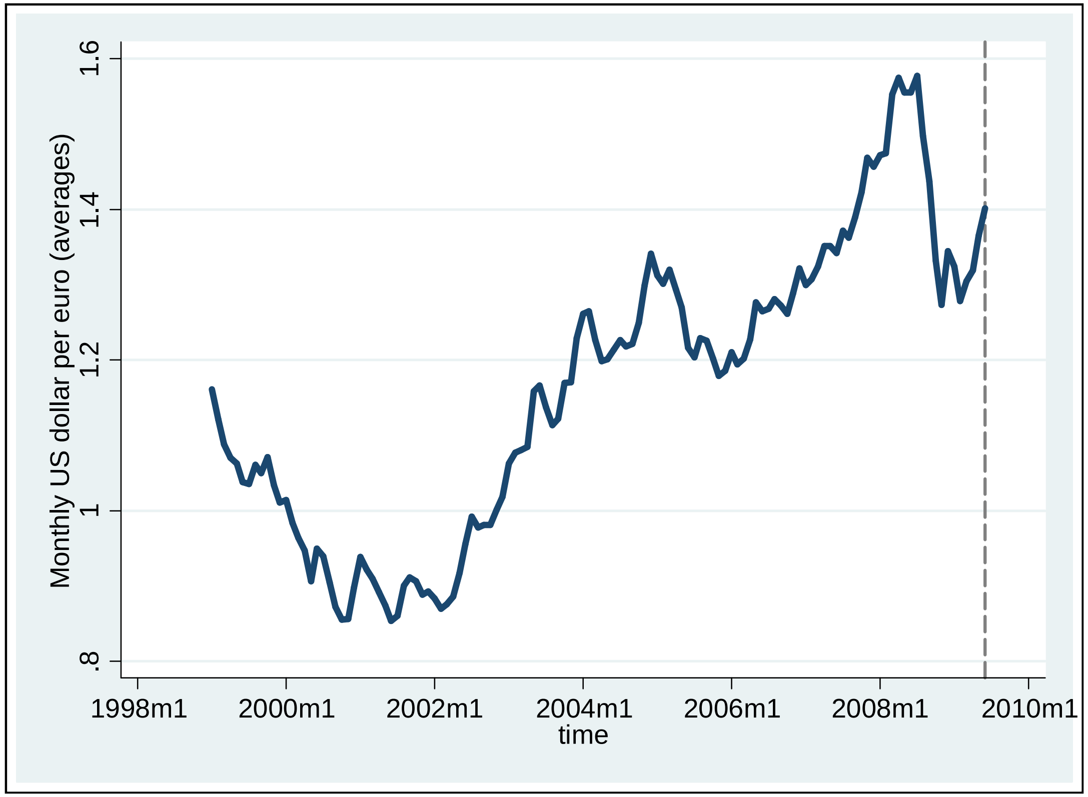


Abbildung 5.8: Entwicklung der schweizerischen, deutschen/europäischen und amerikanischen Euromarktsätze, Monatsdaten, Januar 1983 – Dezember 2002

Nonstationarity: Trending ?

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Example: Nonstationarity AR

AR with φ_1=1

Random walk with drift $y_t = y_{t-1} + \delta + \varepsilon_t$

$$y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

$$E(y_t) = E\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = y_0 + \delta t$$

In contrast, a random walk (δ = 0) is mean stationary.

$$Var(y_t) = Var\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = t \cdot \sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})] = (t-k) \cdot \sigma_\varepsilon^2$$

Nonstationarity

Nonstationary **not** due to trending could be handled by an ARMA model with time-changing parameters:

$$y_t = a(t) y_{t-1} + b(t) \varepsilon_t$$

… but we will focus on models with time invariant parameters.

Notation

Discussing stationarity conditions of ARMA models and extending them to nonstationary series (ARIMA models) is facilitated by introducing ‘time-series notation:

- (polynomials in the) **lag operator *L***
- (first-) **difference operator *Δ***

ARIMA(p,d,q):

$$\underbrace{a(L)(1-L)^d Y_t}_{\underbrace{\Delta^d Y_t}_{X_t}} = \delta + b(L)\varepsilon_t$$

$$\underbrace{X_t - \varphi_1 X_{t-1} - \ldots - \varphi_p X_{t-p}}_{a(L)X_t} = \delta + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}}_{b(L)\varepsilon_t}$$

So Y_t ~ **ARIMA**(p,d,q) and X_t ~ **ARMA**(p,q).

d-times differencing of Y_t yields stationary ARMA X_t

First Difference Operator *Δ*

“The first difference operator, *Δ*, can be manipulated in a similar way to the lag operator, since $\Delta = 1 - L$. The relationship between the two operators can often be usefully exploited.”

For example:

$$(1-L)y_t = y_t - y_{t-1} = \Delta y_t$$

$$L(1-L)y_t = y_{t-1} - y_{t-2} = \Delta y_{t-1}$$

$$\Delta^2 y_t = (1-L)^2 y_t = (1-2L+L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$$

Harvey (1981) “Time series models”, p.26

Mechanics in Differencing $x_t = \Delta^d y_t = (1-L)^d y_t$

d = 1: $x_t = \Delta y_t = (1-L)y_t = y_t - y_{t-1}$

d = 2: $x_t = \Delta^2 y_t = (1-L)^2 y_t$

$$= \Delta\Delta y_t = \Delta(y_t - y_{t-1}) = \Delta y_t - \Delta y_{t-1}$$

$$= y_t - y_{t-1} - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

$$\neq y_t - y_{t-2}$$

d = 3: $x_t = \Delta^3 y_t = (1-L)^3 y_t = (1-L)^2 (y_t - y_{t-1})$

$$= (1 - 2L + L^2)(y_t - y_{t-1})$$

$$= y_t - 2y_{t-1} + y_{t-2} - (y_{t-1} - 2y_{t-2} + y_{t-3})$$

$$= y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3}$$

Mechanics in Differencing

If $x_t = \Delta^d y_t$ then $y_t = \Sigma^d x_t$

Example: ***d* = 1**

$$x_t = \Delta y_t = (1 - L) y_t = y_t - y_{t-1}$$

$$y_t = y_{t-1} + x_t$$

$$y_{t-1} = y_{t-2} + x_{t-1}$$

$$\Rightarrow y_t = y_{t-2} + x_t + x_{t-1}$$

$$y_t = \sum_{i=0}^{\infty} x_{t-i} = \sum_{i=-\infty}^{t} x_i = \sum_{i=-\infty}^{1} x_i + \sum_{i=2}^{t} x_i = y_1 + x_2 + x_3 + \ldots + x_t$$

y_t is obtained from '**integrating**´ (summing) the changes

Mechanics in Differencing

Example: **$d = 2$**

$$x_t = \Delta^2 y_t = \Delta\Delta y_t \qquad y_t = \Sigma^2 x_t = \Sigma\Sigma x_t = \Sigma\Delta y_t$$

$$\Delta y_t = \sum_{i=-\infty}^{t} x_i = \sum_{i=-\infty}^{1} x_i + \sum_{i=2}^{t} x_i = \Delta y_1 + \sum_{i=2}^{t} x_i$$

$$y_t = \Sigma\Delta y_t = \Sigma\underbrace{\left(\Delta y_1 + x_2 + x_3 + \ldots + x_t\right)}_{z_t} = \sum_{i=-\infty}^{1} z_i + \sum_{i=2}^{t} z_i$$

$$= y_1 + z_2 + z_3 + \ldots + z_t$$

$$= y_1 + \underbrace{\Delta y_1 + x_2}_{z_1} + \underbrace{\Delta y_1 + x_2 + x_3}_{z_2} + \ldots + \underbrace{\Delta y_1 + x_2 + x_3 + \ldots + x_t}_{z_t}$$

The lag operator *L*

Applied to a variable with a time subscript:

$$Ly_t = y_{t-1}$$

$$L^2 y_t = L(Ly_t) = L(y_{t-1}) = y_{t-2}$$

$$L^k y_t = y_{t-k} \qquad k = 1,2,3,\ldots$$

$$L^0 y_t = y_t$$

Applied to a constant C:

$$LC = C$$

$$L^2 C = C$$

$$L^k C = C \qquad k = 1,2,3,\ldots$$

$$\left(\alpha L + \beta L^2 + \gamma L^3\right)\cdot C = (\alpha + \beta + \gamma)\cdot C$$

Johnston/DiNardo (1997) “Econometric Methods”, p. 206

The lag operator *L*

The lag operator follows the same algebraic rules as the multiplication operator ("multiply y_t by L"):

$$L(\beta \cdot y_t) = \beta \cdot L y_t$$
$$L(y_t + w_t) = L y_t + L w_t$$

$$(1 - L) y_t = y_t - y_{t-1}$$
$$L(1 - L) y_t = L y_t - L^2 y_t = y_{t-1} - y_{t-2}$$

Hamilton (1994) "Time Series Analysis", p. 27

Examples for alternative expressions

$$
\begin{aligned}
x_t &= (a + bL)Ly_t \\
&= (aL + bL^2)y_t \\
&= ay_{t-1} + by_{t-2}
\end{aligned}
\qquad
\begin{aligned}
x_t &= (1-\lambda_1 L)(1-\lambda_2 L)y_t \\
&= (1-\lambda_1 L - \lambda_2 L + \lambda_1\lambda_2 L^2)y_t \\
&= [1-(\lambda_1+\lambda_2)L + \lambda_1\lambda_2 L^2]y_t \\
&= y_t - (\lambda_1+\lambda_2)y_{t-1} + \lambda_1\lambda_2 y_{t-2}
\end{aligned}
$$

"An expression such as ($aL + bL^2$) is called **polynomial in the lag operator**.

It is algebraically similar to a simple polynomial ($az + bz^2$) where z is a scalar. The difference is that the simple polynomial refers to a particular number, whereas a polynomial in the lag operator refers to an operator that would be applied to one time series to produce a new time series."

Hamilton (1994) "Time Series Analysis", p. 27

AR processes in lag operator notation

AR(1)

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

$$y_t = \varphi L y_t + \varepsilon_t$$

$$(1 - \varphi L) y_t = \varepsilon_t$$

$$a_1(L) y_t = \varepsilon_t$$

AR(2)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

$$y_t = \varphi_1 L y_t + \varphi_2 L^2 y_t + \varepsilon_t$$

$$\left(1 - \varphi_1 L - \varphi_2 L^2\right) y_t = \varepsilon_t$$

$$a_2(L) y_t = \varepsilon_t$$

AR(*p*)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t$$

$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p\right) y_t = \varepsilon_t$$

$$a_p(L) y_t = \varepsilon_t$$

MA processes in lag operator notation

MA(1)

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1}$$
$$y_t = \varepsilon_t - \theta L\varepsilon_t$$
$$y_t = (1-\theta L)\varepsilon_t$$
$$y_t = b_1(L)\varepsilon_t$$

MA(2)

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$
$$y_t = \varepsilon_t - \theta_1 L\varepsilon_t - \theta_2 L\varepsilon_t$$
$$y_t = \left(1-\theta_1 L - \theta_2 L^2\right)\varepsilon_t$$
$$y_t = b_2(L)\varepsilon_t$$

MA(*q*)

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$
$$y_t = \left(1-\theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right)\varepsilon_t$$
$$y_t = b_q(L)\varepsilon_t$$

ARMA processes in lag operator notation

ARMA(1,1)

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$(1 - \varphi_1 L) y_t = (1 - \theta_1 L) \varepsilon_t$$

$$a_1(L) y_t = b_1(L) \varepsilon_t$$

ARMA(*p*,*q*)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p\right) y_t = \left(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right) \varepsilon_t$$

$$a_p(L) y_t = b_q(L) \varepsilon_t$$

Inverse of a lag-operator polynomial

First-order polynomial $1-\alpha L = A(L)$

Multiply by the following operator: $\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots+\alpha^pL^p\right)$

$$\begin{aligned}
&(1-\alpha L)\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots+\alpha^pL^p\right)\\
&=\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots+\alpha^pL^p\right)-\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots+\alpha^pL^p\right)\alpha L\\
&=\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots+\alpha^pL^p\right)-\alpha L-\alpha^2L^2-\alpha^3L^3-\ldots-\alpha^pL^p-\alpha^{p+1}L^{p+1}\\
&=\left(1-\alpha^{p+1}L^{p+1}\right)
\end{aligned}$$

Provided that $|\alpha| < 1$: $\lim\limits_{p\to\infty}\left(\alpha^{p+1}L^{p+1}\right)=0$

Hence, as $p \to \infty$: $\quad (1-\alpha L)\left(1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots\right)=1$

$$\Rightarrow A^{-1}(L)=\frac{1}{(1-\alpha L)}=1+\alpha L+\alpha^2L^2+\alpha^3L^3+\ldots$$

Johnston/DiNardo (1997) “Econometric Methods”, p. 206

Example: First-order autoregressive process AR(1)

$$y_t = \varphi \cdot y_{t-1} + \varepsilon_t \Leftrightarrow (1-\varphi L)y_t = \varepsilon_t \quad \Leftrightarrow a(L)y_t = \varepsilon_t$$

If $|\varphi| < 1$:

$$y_t = \frac{\varepsilon_t}{(1-\varphi L)} = \left(1 + \varphi L + \varphi^2 L^2 + \varphi^3 L^3 + \ldots\right)\varepsilon_t = \sum_{j=0}^{\infty} (\varphi L)^j \varepsilon_t$$

$$= \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \ldots$$

$$= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

$$= \left(1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \ldots\right)\varepsilon_t = c(L)\varepsilon_t$$

→ A **stationary AR(1)** can be written as an **infinite MA**. with lag polynomial $c(L)$.

Example: First-order moving average process MA(1)

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 L)\varepsilon_t = b(L)\varepsilon_t$$

If $|\theta_1| < 1$:

$$\varepsilon_t = \frac{y_t}{(1-\theta_1 L)} = (1 + \theta_1 L + \theta_1^2 L^2 + \theta_1^3 L^3 + \ldots) y_t = \sum_{j=0}^{\infty} (\theta L)^j y_t$$

$$= y_t + \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \ldots$$

$$= y_t + g_1 y_{t-1} + g_2 y_{t-2} + \cdots$$

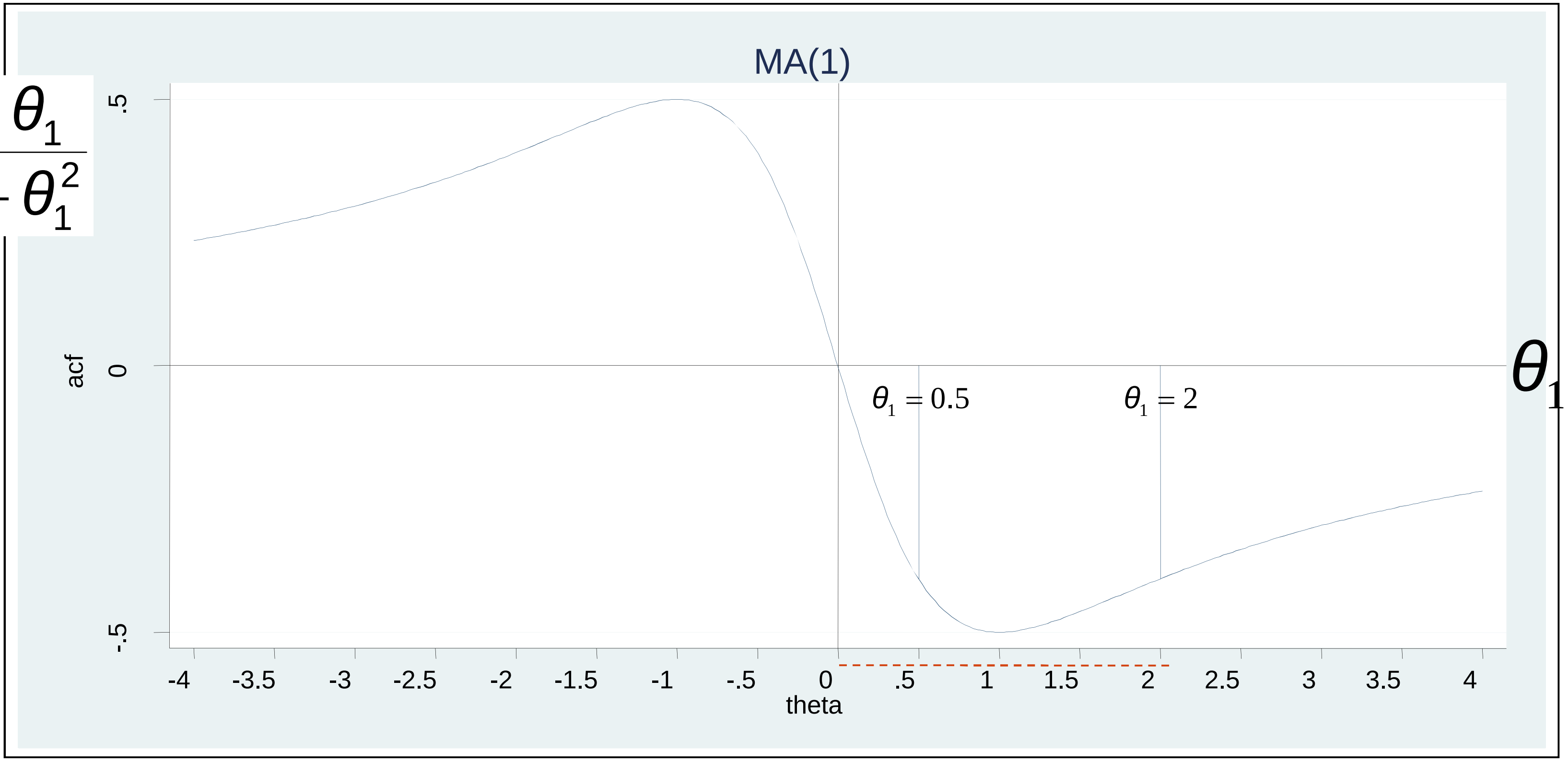
$$= (1 + g_1 L + g_2 L^2 + g_3 L^3 + \ldots) y_t = d(L) y_t$$

An **MA(1)** that can be written as an **infinite AR** process with lag operator polynomial $d(L)$ is called “invertible”.

Recall: Moving Average Models

ACF of MA(1) $y_t = \varepsilon_t - 0.5\,\varepsilon_{t-1}$ vs. $y_t = \varepsilon_t - 2\,\varepsilon_{t-1}$

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$



Both have same ACF (same serial correlation) but only the process with θ_1=0.5 is invertible.

Lag Operator

stationary AR and invertible MA are “mirror images”

AR(1): If $|\varphi_1| < 1$:

$$y_t = \varphi_1 \cdot y_{t-1} + \varepsilon_t$$
$$(1-\varphi_1 L) y_t = \varepsilon_t$$
$$a(L) y_t = \varepsilon_t$$
$$y_t = a^{-1}(L)\, \varepsilon_t$$
$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$
$$y_t = c(L)\, \varepsilon_t$$
$$\rightarrow \mathrm{MA}(\infty)$$

MA (1): If $|\theta_1| < 1$:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$
$$y_t = (1-\theta_1 L)\varepsilon_t$$
$$y_t = b(L)\varepsilon_t$$
$$b^{-1}(L)\, y_t = \varepsilon_t$$
$$\varepsilon_t = y_t + g_1 y_{t-1} + g_2 y_{t-2} + \cdots$$
$$\varepsilon_t = d(L) y_t$$
$$\rightarrow \mathrm{AR}(\infty)$$

Process	acf	pacf
AR	Tails off toward zero	Cuts off to zero (after lag p)
MA	Cuts off to zero (after lag *q*)	Tails off toward zero

Conditions for Stationarity

Example: AR(2) $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$

$$\underbrace{\left(1-\varphi_1 L - \varphi_2 L^2\right)}_{\text{AR polynomial}} y_t = \varepsilon_t$$

Stationary conditions of AR(2) can be stated via

1. roots of the **lag order polynomial**

   $$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 = 0$$

   stationary if : $|z_1| > 1$ and $|z_2| > 1$

   "z_s outside the unit circle"

2. roots of the **characteristic equation**

   $$\lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

   stationary if : $|\lambda_1| < 1$ and $|\lambda_2| < 1$

   "λ_s inside the unit circle"

Conditions for Stationarity

How do we find these roots? Solve quadratic equation!

For the **lag operator polynomial** $1-\varphi_1 z-\varphi_2 z^2=0$

$$z_1, z_2=\frac{\varphi_1 \pm \sqrt{\varphi_1^2+4\varphi_2}}{-2\varphi_2}$$

Note: If $\sqrt{\varphi_1^2+4\varphi_2}<0$ then roots are complex.

Example: $\left(1-0.6L+0.08L^2\right)y_t=\varepsilon_t$ i.e. $\varphi_1=0.6$ $\varphi_2=-0.08$

$$z_1=\frac{0.6+\sqrt{0.6^2+4(-0.08)}}{-2\cdot(-0.08)}=5 \qquad z_2=\frac{0.6-\sqrt{0.6^2+4(-0.08)}}{-2\cdot(-0.08)}=2.5$$

$|z_1|>1$ and $|z_1|>1$ $\rightarrow$ the process is stationary

Conditions for Stationarity

How do we find these roots?

For the **characteristic equation** $\lambda^2 - \varphi_1\lambda - \varphi_2 = 0$

$$\lambda_1, \lambda_2 = \frac{\varphi_1 \pm \sqrt{\varphi_1^2 + 4\varphi_2}}{2}$$

Note: If $\sqrt{\varphi_1^2 + 4\varphi_2} < 0$
then λ_1 and λ_2 are complex.

Example: $\varphi_1 = 0.6 \quad \varphi_2 = -0.08 \quad \lambda^2 - 0.6\lambda + 0.08 = 0$

$$a_1 = \frac{0.6 - \sqrt{0.6^2 + 4(-0.08)}}{2} = 0.2 \qquad a_2 = \frac{0.6 + \sqrt{0.6^2 + 4(-0.08)}}{2} = 0.4$$

$|\lambda_1|<1$ and $|\lambda_2|<1 \rightarrow$ the process is stationary

Conditions for Stationarity

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 = 0$$

$$\lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

Are these roots related? Yes! $\Rightarrow z_j = \dfrac{1}{\lambda_j} \quad j = 1, 2$

Roots of the lag polynomial and roots of the characteristic equation are reciprocals!

Example: $(1 - 0.6L + 0.08L^2) y_t = \varepsilon_t$

$z_1 = 5 \quad \Leftrightarrow \quad a_1 = 0.2 = \dfrac{1}{5}$ $\qquad z_2 = 2.5 \quad \Leftrightarrow \quad a_2 = 0.4 = \dfrac{1}{2.5}$

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 = 0 \qquad \lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

$$z_j = \frac{1}{\lambda_j} \qquad j = 1, 2$$

To show this, factor a(z)!

$$1 - \varphi_1 z - \varphi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

$$\text{Note}: \quad 1 - (\lambda_1 + \lambda_2) z + \lambda_1 \lambda_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

Example: $(1 - 0.6L + 0.08L^2) y_t = \varepsilon_t$

$$(1 - 0.6z + 0.08z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

$$(1 - 0.6z + 0.08z^2) = (1 - 0.2z)(1 - 0.4z)$$

Conditions for Stationarity

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 = 0 \qquad\qquad \lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

$$z_j = \frac{1}{\lambda_j} \qquad j = 1, 2$$

Factoring a(z): $1 - \varphi_1 z - \varphi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z)$

For the roots $\quad z_j = \dfrac{1}{\lambda_j} \qquad j = 1, 2$

Example: $\quad (1 - 0.6L + 0.08L^2) y_t = \varepsilon_t$

$$(1 - 0.6z + 0.08z^2) = (1 - 0.2z)(1 - 0.4z)$$

$$z_1 = 5 \quad \Leftrightarrow \quad a_1 = 0.2 = \frac{1}{5} \qquad\qquad z_2 = 2.5 \quad \Leftrightarrow \quad a_2 = 0.4 = \frac{1}{2.5}$$

Conditions for Stationarity

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 = 0 \qquad \lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

$$z_j = \frac{1}{\lambda_j} \qquad j = 1, 2$$

To show this, factor a(z)!

$$1 - \varphi_1 z - \varphi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

$$z^{-2} - \varphi_1 z^{-1} - \varphi_2 = (z^{-1} - \lambda_1)(z^{-1} - \lambda_2)$$

divide by z^2

$$\lambda^2 - \varphi_1 \lambda - \varphi_2 = (\lambda - \lambda_1)(\lambda - \lambda_2)$$

define $\lambda = z^{-1}$

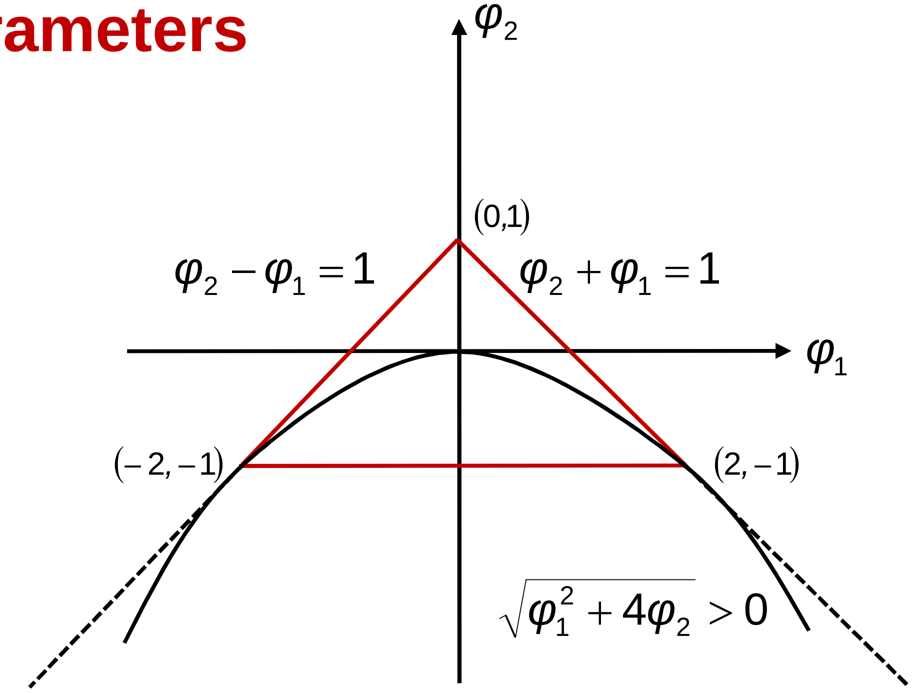
Hence, the roots os the charactersitic equation are indeed equal to the roots of the factorized a(z)!

In terms of the AR-parameters

$|\lambda_1| < 1$ and $|\lambda_2| < 1$

In terms of φ:

$$\varphi_2 + \varphi_1 < 1$$
$$\varphi_2 - \varphi_1 < 1$$
$$\varphi_2 > -1$$



Stationarity whenever (φ_2, φ_1) are inside the triangle.

Hamilton (1994) "Time Series Analysis", p. 17/18

Nonstationarity

Example: $$y_t = 174 y_{t-1} - 0.74 y_{t-2} + 0.021 + \varepsilon_t$$

$$y_t - 174 y_{t-1} + 0.74 y_{t-2} = 0.021 + \varepsilon_t$$
$$(1 - 174L + 0.74L^2) y_t = 0.021 + \varepsilon_t$$

$$y_t - 0.74 y_{t-1} - y_{t-1} + 0.74 y_{t-2} = 0.021 + \varepsilon_t$$
$$y_t - y_{t-1} - 0.74(y_{t-1} - y_{t-2}) = 0.021 + \varepsilon_t$$
$$(1 - 0.74L)(y_t - y_{t-1}) = 0.021 + \varepsilon_t$$
$$(1 - 0.74L)(1 - L) y_t = 0.021 + \varepsilon_t$$

AR polynomial has a root of $z = 1$ ("**unit root**").
→ The process is nonstationary.

Example:

$$y_t = 1.74 y_{t-1} - 0.74 y_{t-2} + 0.021 + \varepsilon_t$$

$$\underbrace{(1 - 1.74L + 0.74L^2)}_{A(L)} y_t = 0.021 + \varepsilon_t$$

$$(1 - 0.74L)(1 - L) y_t = 0.021 + \varepsilon_t$$

"**unit root**"

$$\underbrace{(1 - 0.74L)}_{a(L)} \Delta y_t = 0.021 + \varepsilon_t$$

y_t is nonstationary but $y_t - y_{t-1} = \Delta y_t = (1 - L)y_t$ is stationary (it has a stationary AR polynomial $a(L)$).

→ Differencing once leads to a stationary AR

Nonstationarity

Example:

$$y_t = 174 y_{t-1} - 0.74 y_{t-2} + 0.021 + \varepsilon_t$$

$$\underbrace{(1 - 174L + 0.74L^2)}_{A(L)} y_t = 0.021 + \varepsilon_t$$

$$\underbrace{(1 - 0.74L)}_{a(L)} \Delta y_t = 0.021 + \varepsilon_t$$

"ARIMA(1,1,0)"

In general, ARIMA(p,d,q)

$$\underbrace{a(L)(1-L)^d}_{A(L)} y_t = \delta + b(L)\varepsilon_t$$

Conditions for Stationarity

Stationary conditions in terms of roots carry over to AR(p)

1. roots of the **lag order polynomial**

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0$$

stationary if for all roots : $|z_j| > 1$
"z_s outside the unit circle"

2. roots of the **characteristic equation**

$$\lambda^p - \varphi_1 \lambda^{p-1} - \varphi_2 \lambda^{p-2} - \ldots - \varphi_p = 0$$

stationary if for all roots : $|\lambda_j| < 1$
"λ_s inside the unit circle"

Non-stationary case: particular attention will be given to the existence of (a) **unit root**(s)

Conditions for Stationarity

Unit root and **AR lag order polynomial**

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0$$

$$a(1) = 1 - \varphi_1 1 - \varphi_2 1^2 - \ldots - \varphi_p 1^p = 0$$

$$\Rightarrow 1 = \varphi_1 + \varphi_2 + \ldots + \varphi_p$$

Example: Null hypotheses of **unit root tests**

$y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t$
$H_0: \varphi_1 = 1$	$H_0: \varphi_1 + \varphi_2 = 1$	$H_0: \varphi_1 + \varphi_2 + \ldots + \varphi_p = 1$

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • **Nonstationarity and Trends** • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Nonstationarity

- Trends are the most obvious and maybe most common form of nonstationarity.
- Two models which allow for trends:
  - **Trend-stationary (TS) models**

    Y_t = deterministic trend + stationary ARMA
  - **Difference-stationary (DS) models**

    ARIMA model with a unit root and a constant term

Examples:

Trend-stationary (TS) model

$$y_t = \delta_0 + \delta_1 t + u_t \quad \text{with} \quad u_t = \varphi_1 u_{t-1} + \varepsilon_t$$

Difference-stationary (DS) model

$$a(L)(1-L)y_t = \delta + b(L)\varepsilon_t \quad \text{with } a(L) = 1 \text{ and } b(L) = 1$$

$$y_t = y_{t-1} + \delta + \varepsilon_t \;\Rightarrow y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

They look similar but can have very different properties

Trend-stationary (TS) models

Y_t = deterministic trend + stationary ARMA

This can be written as: $y_t = \sum_{j=0}^{m} \delta_j \cdot t^j + u_t$

Where u_t is a stationary and invertible ARMA(p,q) process with $E(u_t) = 0$: $a(L)u_t = b(L)\varepsilon_t$

The Uncertain Unit Root in Real GNP

By Glenn D. Rudebusch*

American Economic Review, Vol. 83(1), p 264-72

„Indeed, the common practice of macroeconomists of all theoretical persuasions was to model movements in real GNP as stationary fluctuations around a linear deterministic trend (e.g., Finn Kydland and Edward C. Prescott, 1980; Olivier J. Blanchard, 1981). Such a trendstationary (TS) model of real GNP was the canonical empirical representation of aggregate output until the early 1980's."

Nonstationarity and Trends

$$y_t = \sum_{j=0}^{m} \delta_j \cdot t^j + u_t \qquad a(L)u_t = b(L)\varepsilon_t$$

$$y_t = \delta_0 + \delta_1 t + u_t \quad \text{with} \quad u_t = \varphi_1 u_{t-1} + \varepsilon_t$$

„ The impulse and propagation mechanisms of business cycles have long been debated; however, until recently, economists were in fairly broad agreement that business fluctuations could be studied separately from the secular growth of the economy. This separation was justified because, to a first approximation, the factors underlying trend growth were assumed to be stable at business-cycle frequencies..”

Trend-stationary (TS) models $y_t = \sum_{j=0}^{m} \delta_j \cdot t^j + u_t$

$$E(y_t) = E\left(\sum_{j=0}^{m} \delta_j \cdot t^j + u_t\right) = E\left(\sum_{j=0}^{m} \delta_j \cdot t^j\right) + E(u_t)$$

$$= E\left(\sum_{j=0}^{m} \delta_j \cdot t^j\right) = \sum_{j=0}^{m} \delta_j \cdot t^j = \mu_t$$

The mean is not independent of time.

$$Cov(y_t, y_{t+k}) = E[(y_t - \mu_t)(y_{t+k} - \mu_{t+k})] = E(u_t u_{t+k}) = \gamma_k$$

The covariance is independent of time.

Trend-stationary (TS) models $y_t = \sum_{j=0}^{m} \delta_j \cdot t^j + u_t$

Example: $y_t = \delta_0 + \delta_1 t + u_t$

$$u_t = \varphi_1 u_{t-1} + \varepsilon_t$$

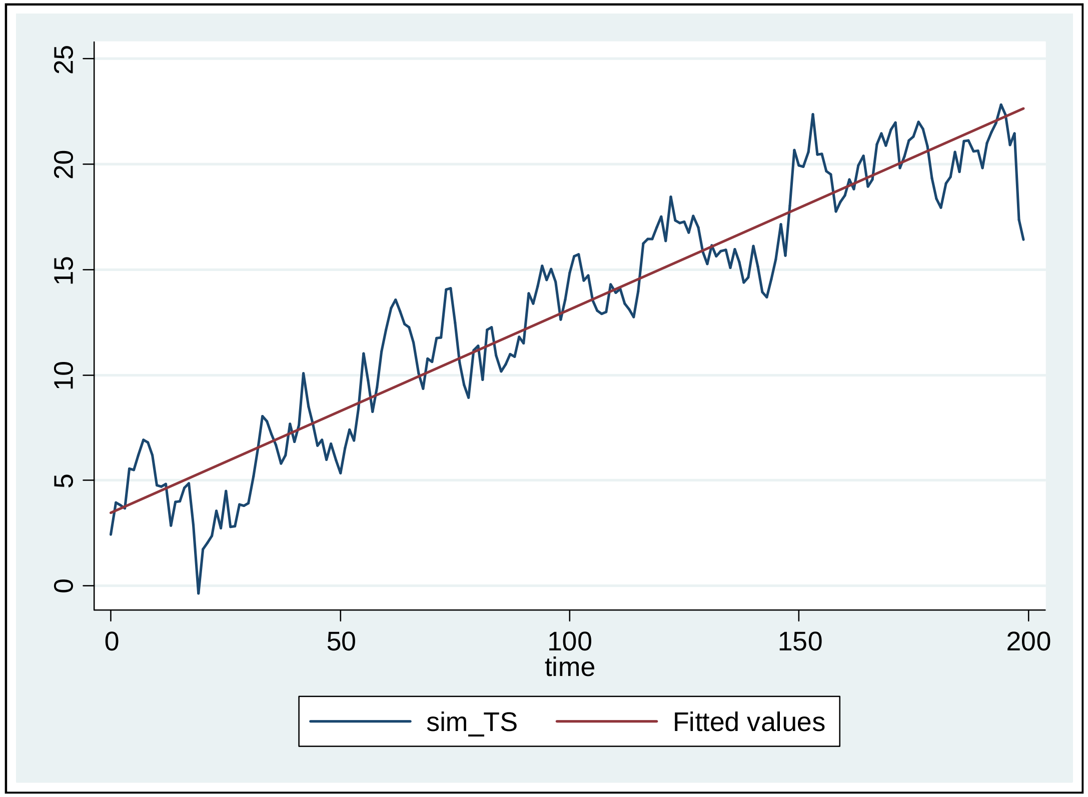
If we remove the trend, we have a stationary series:

$$y_t - \delta_0 - \delta_1 t = u_t$$

If $\delta_1 > 0$ then the series will fluctuate around an upward trend „but with no obviously tendency for the amplitude of the fluctuations to increase or decrease.“

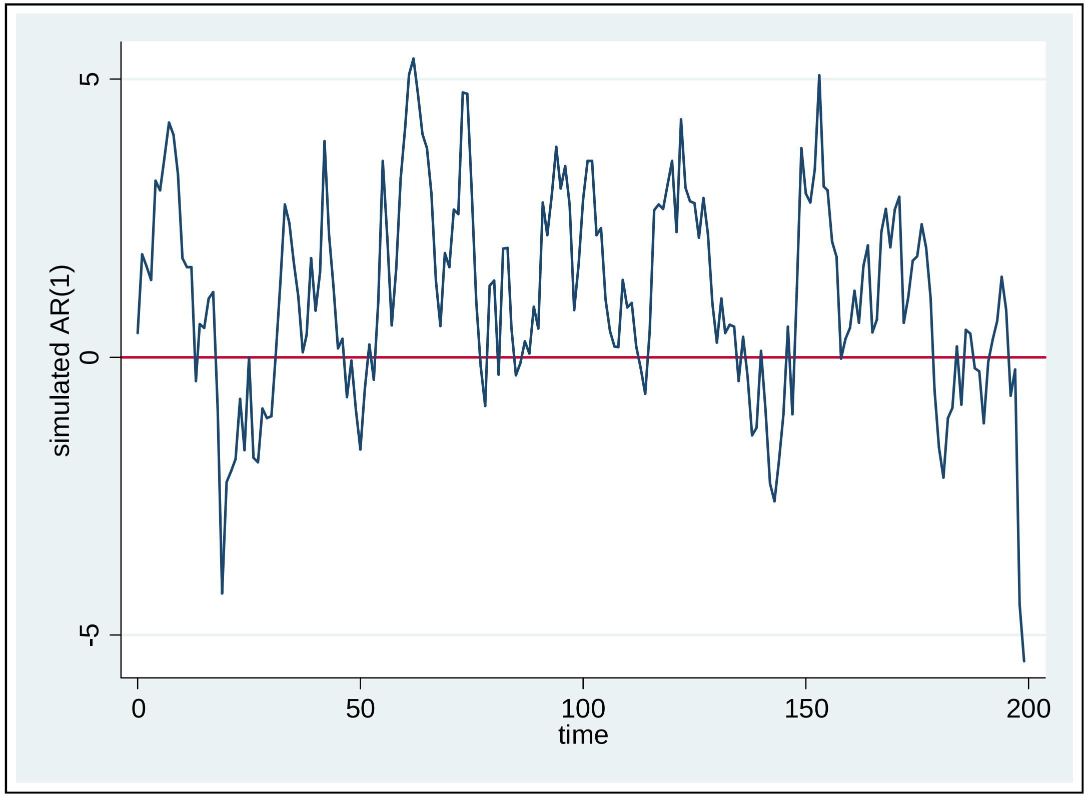
Example 1:

$$y_t = \delta_0 + \delta_1 t + u_t = 2 + 0.1t + u_t$$
$$u_t = \varphi_1 u_{t-1} + \varepsilon_t = 0.9u_{t-1} + \varepsilon_t$$



Verbeek (2000) "A Guide to Modern Econometrics "

Example 1: $y_t = \varphi_1 y_{t-1} + \varepsilon_t = 0.9 y_{t-1} + \varepsilon_t$

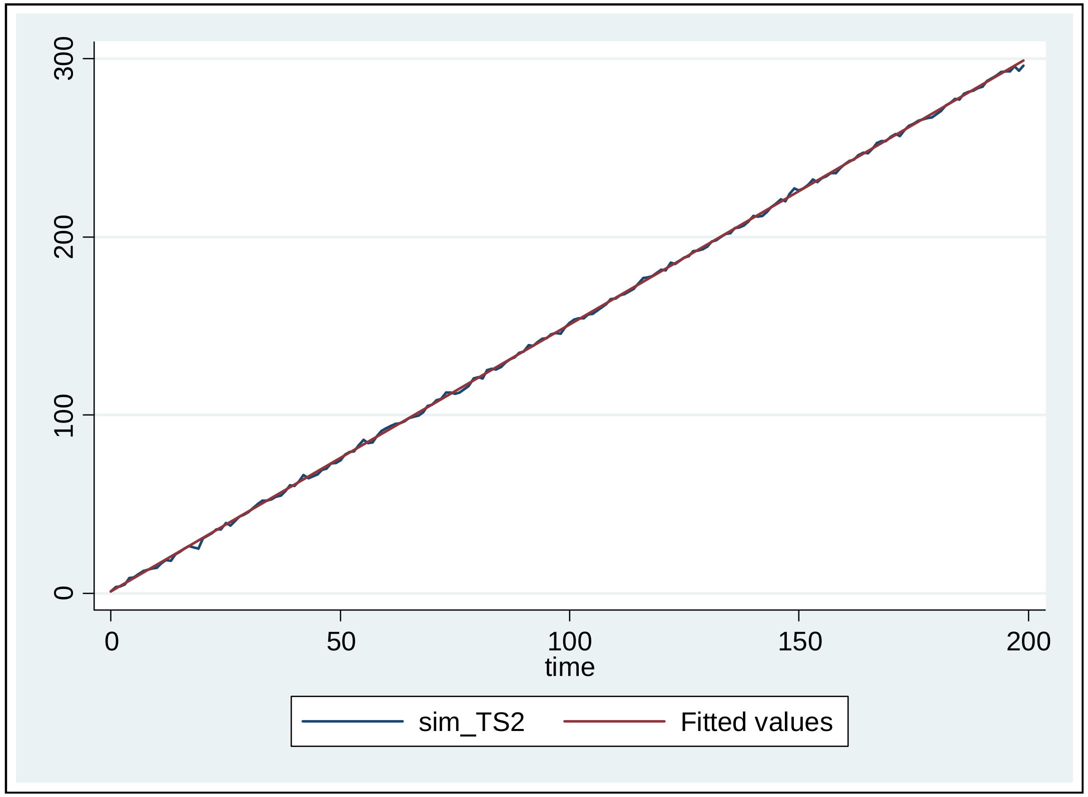


Verbeek (2000) “A Guide to Modern Econometrics ”

Example 2:

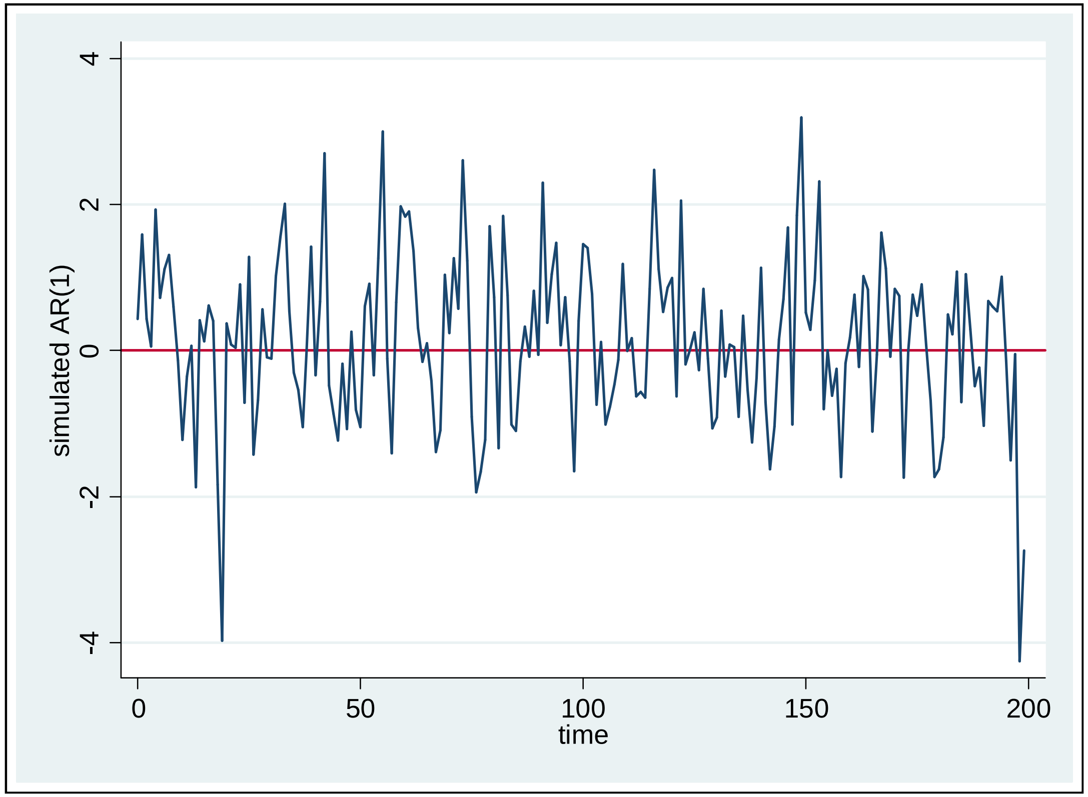
$$y_t = \delta_0 + \delta_1 t + u_t = 0.5 + 15t + u_t$$

$$u_t = \varphi_1 u_{t-1} + \varepsilon_t = 0.3u_{t-1} + \varepsilon_t$$



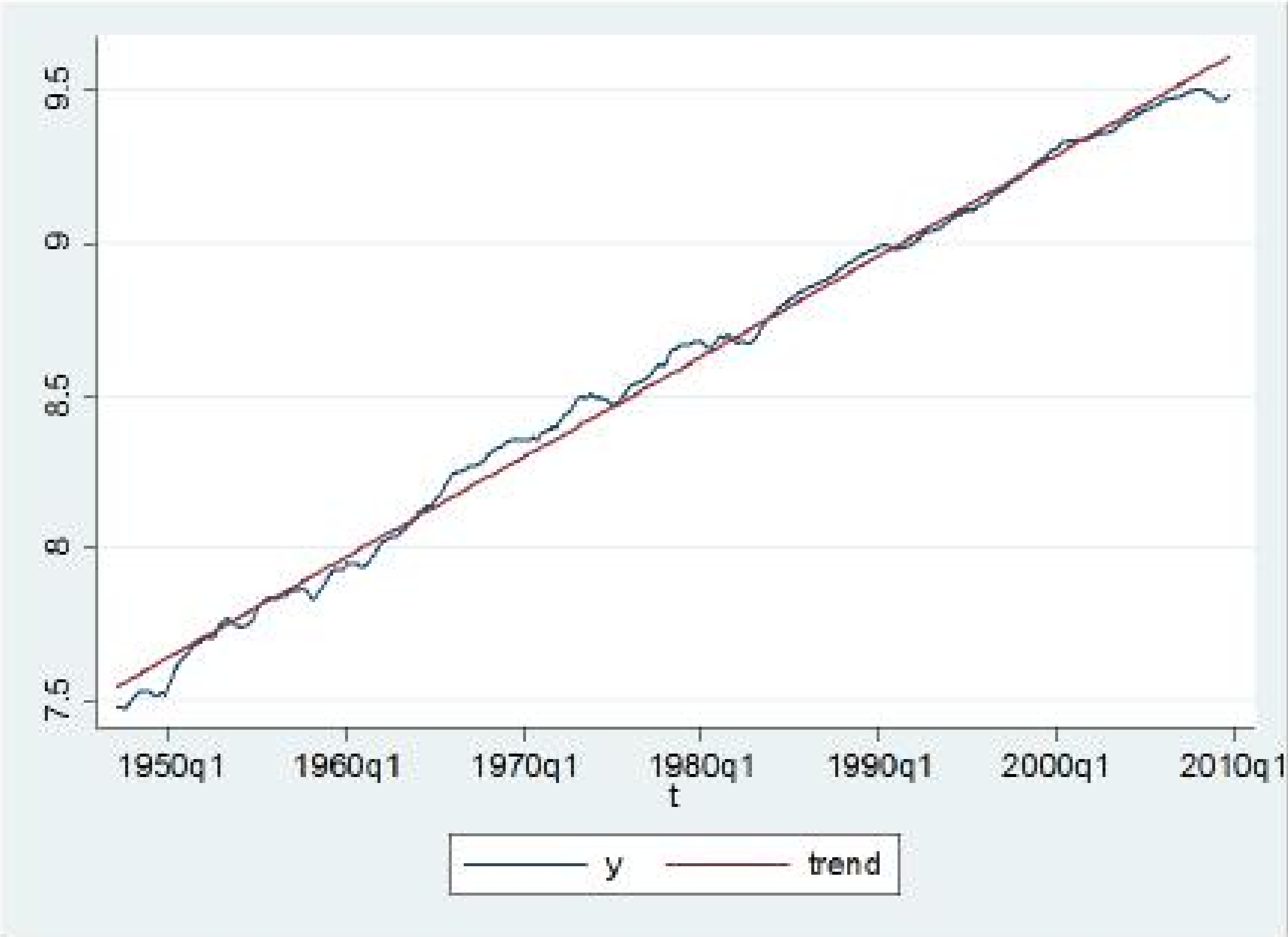
Verbeek (2000) “A Guide to Modern Econometrics ”

Example 2: $y_t = \varphi_1 y_{t-1} + \varepsilon_t = 0.3 y_{t-1} + \varepsilon_t$



Verbeek (2000) “A Guide to Modern Econometrics ”

log of real GDP



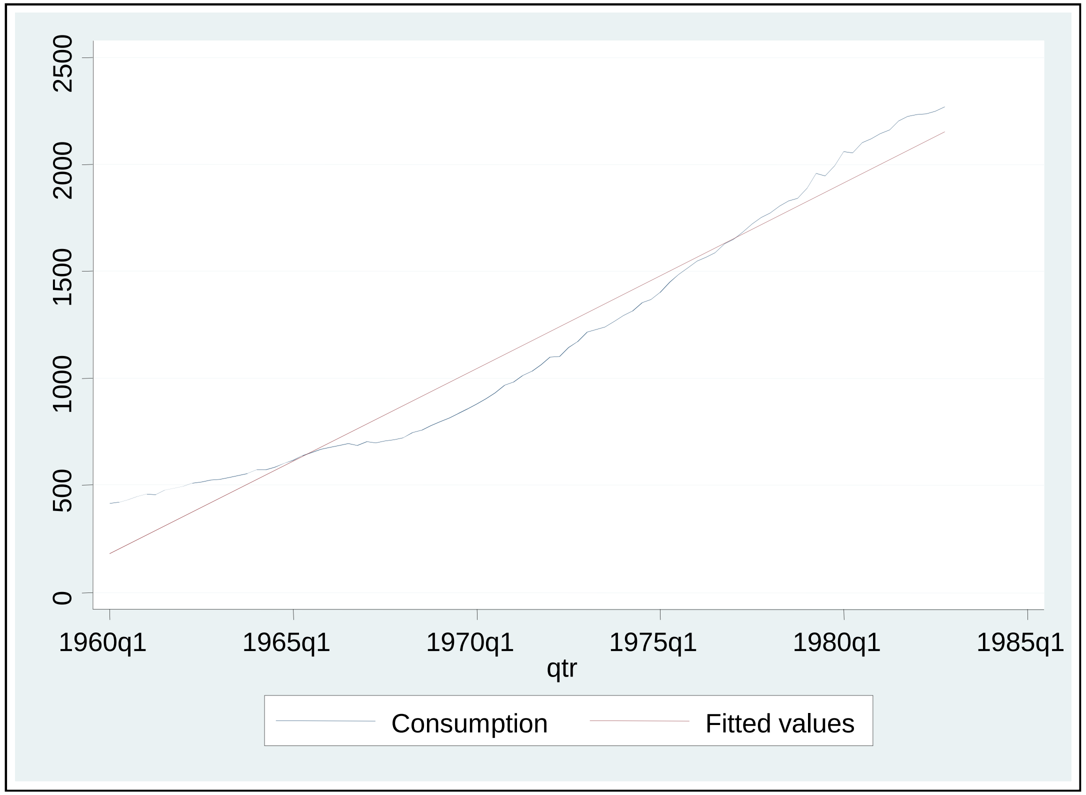
Excursus: Logarithmic Transformation

Before we even remove a deterministic trend in the TS model or difference in the DS model it is often useful to first take logs of the original series.

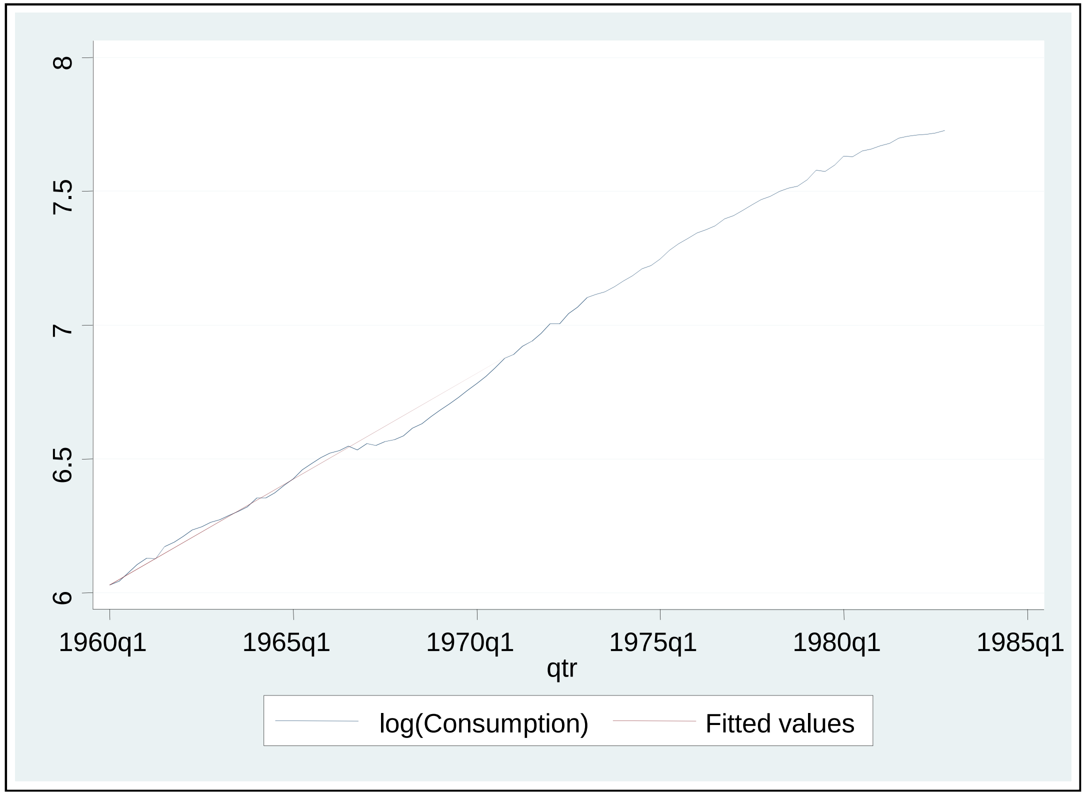
This will linearize an exponential trend, i.e. constant proportional growth.

$$\log\left(e^{\delta t}\right) = \delta t$$

Example: Consumption in Germany



Example: log(Consumption) in Germany



Excursus: Logarithmic Transformation

Moreover, 1st differences of log series are approximately growth rates (percentage changes) which can be expected to be stationary even if the original series is not.

$$\log(y_t) - \log(y_{t-1}) = (1-L)\log(y_t) = \log\left(\frac{y_t}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

Differences of logs can be interpreted as being proportional to the percentage change in the original variable:

$$y_t = (1+p_t)y_{t-1}$$

$$\log(y_t) - \log(y_{t-1}) = (1-L)\log(y_t) = \ln(1+p_t) \approx p_t$$

The data consist of quarterly observations on U.S. postwar log real GNP per capita from 1948:3 to 1988:4.

$$\text{(1)} \quad Y_t = \underset{(0.109)}{-0.321} + \underset{(0.00010)}{0.00030t} + \underset{(0.073)}{1.335Y_{t-1}}$$

$$\underset{(0.073)}{-0.401Y_{t-2}} + u_t \qquad \hat{\sigma}_u = 0.01013$$

(standard errors of the coefficients appear in parentheses).[4] I will refer to this specific model estimate for the sample as the TS_{OLS} model.

$$y_t = -0.321 + 0.0003t + 1.335y_{t-1} - 0.401y_{t-1} + \varepsilon_t$$

$$y_t = \sum_{j=0}^{m} \delta_j \cdot t^j + u_t \qquad a(L)u_t = b(L)\varepsilon_t$$

$$y_t = \delta_0 + \delta_1 t + u_t \qquad u_t = \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \varepsilon_t$$

$$y_t = \delta_0 + \delta_1 t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \varepsilon_t$$

$$y_{t-1} = \delta_0 + \delta_1(t-1) + u_{t-1} \Rightarrow u_{t-1} = y_{t-1} - \delta_0 - \delta_1(t-1)$$

$$y_{t-2} = \delta_0 + \delta_1(t-2) + u_{t-2} \Rightarrow u_{t-2} = y_{t-2} - \delta_0 - \delta_1(t-2)$$

$$y_t = \underbrace{\left[(1-\varphi_1-\varphi_2)\delta_0 + (\varphi_1 + 2\varphi_2)\delta_1\right]}_{\delta} + \underbrace{\left[(1-\varphi_1-\varphi_2)\delta_1\right]}_{\gamma} t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

"In contrast to previous work, much of the research of the last ten years has assumed a unit root in the autoregressive representation of real GNP, which is inconsistent with a TS model of output. A model with a unit root (is) commonly termed a "difference stationary" (DS) model."

$$a(L)(1-L)y_t = \delta + b(L)\varepsilon_t$$

Difference-stationary (DS) models

ARIMA model with a unit root (and a constant term)

$$a(L)(1-L)y_t = \delta + b(L)\varepsilon_t$$

Special cases: AR(1) with $\varphi_1 = 1$:

Random walk $y_t = y_{t-1} + \varepsilon_t$

Random walk with drift $y_t = y_{t-1} + \delta + \varepsilon_t$

Repeated substitution yields:

$$y_t = y_0 + \underbrace{\delta t}_{\text{"built-in deterministic trend"}} + \underbrace{\sum_{j=1}^{t} \varepsilon_j}_{\text{all past shocks influence } y_t}$$

Application: The Uncertain Unit Root in Real GNP

"A model with a unit root, commonly termed a "difference stationary" (DS) model, implies that any stochastic shock to output contains an element that represents a permanent shift in the level of the series. If real GNP is best represented by a DS model, the traditional separation between business cycles and trend growth is incorrect."

$$y_t = y_0 + \underbrace{\delta t}_{\text{"built-in deterministic trend"}} + \underbrace{\sum_{j=1}^{t} \varepsilon_j}_{\text{all past shocks influence } y_t}$$

Difference-stationary (DS) models

Random walk with drift $y_t = y_{t-1} + \delta + \varepsilon_t$

$$y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

$$E(y_t) = E\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = y_0 + \delta t$$

In contrast, a random walk ($\delta = 0$) is mean stationary.

$$Var(y_t) = Var\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = t \cdot \sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})] = (t-k) \cdot \sigma_\varepsilon^2$$

Unit root processes and real world

Ericsson: Comment on "Economic Forecasting in a Changing World"

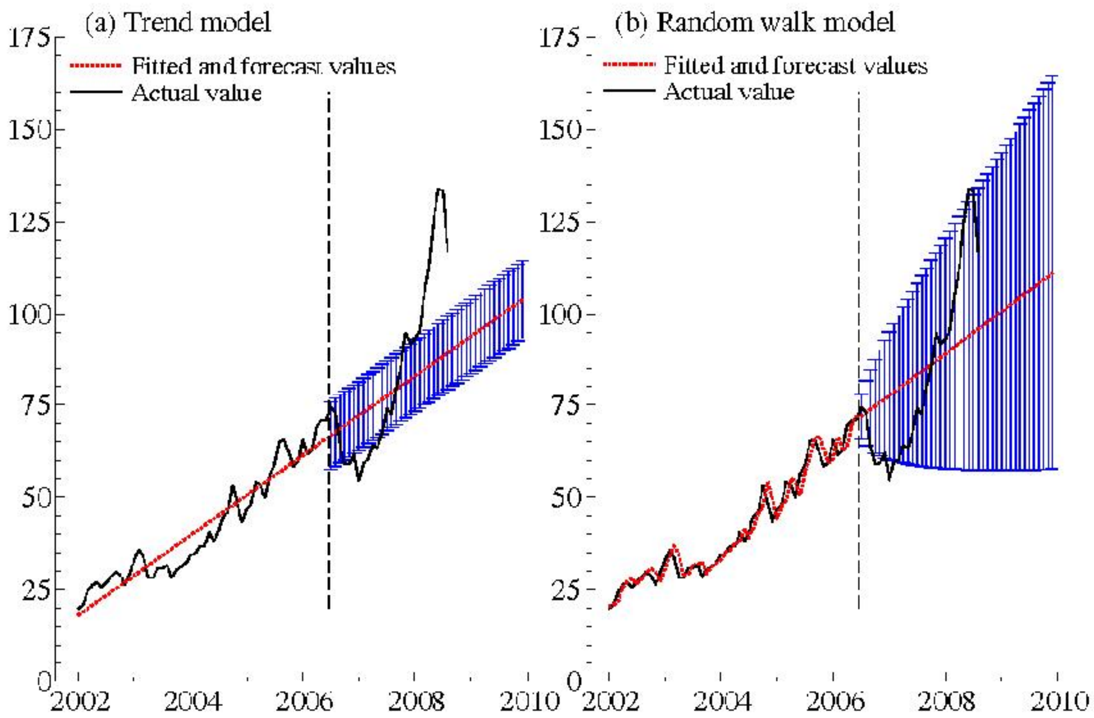


Figure 2: Actual, fitted, and forecast values from the trend and random walk models of the oil price, with 95% confidence intervals for the forecasts.

Stationary vs. nonstationary AR(1)

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t$$

Variance and covariance if $|\varphi| < 1$:

$$Var(y_t) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2} \qquad Cov(y_t, y_{t-1}) = \gamma_1 = \varphi_1 \gamma_0 = \frac{\varphi_1 \sigma_\varepsilon^2}{1-\varphi_1^2}$$

Variance and covariance for a fixed y_0 and $\varphi = 1$:

$$Var(y_t) = \gamma_0 = t \cdot \sigma_\varepsilon^2$$

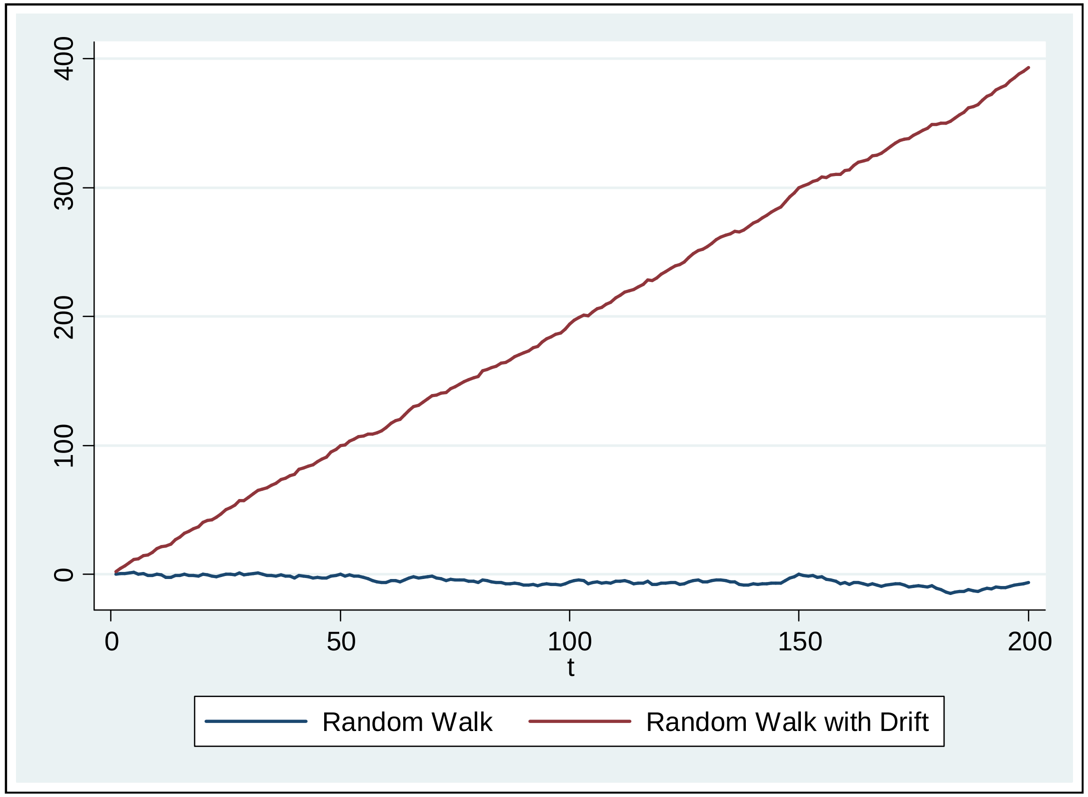
$$Cov(y_t, y_{t-1}) = \gamma_1 = (t-1) \cdot \sigma_\varepsilon^2$$

Johnston, DiNardo (1997) “Econometric Methods”

Example:

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t = y_{t-1} + \varepsilon_t$$

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t = y_{t-1} + 2 + \varepsilon_t$$



Difference-stationary (DS) model

$$y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

In contrast: Trend-stationary (TS) model

$$y_t = \delta_0 + \delta_1 t + u_t \quad \text{with} \quad u_t = \varphi_1 u_{t-1} + \varepsilon_t$$

Looks similar (both DS and TS model include a deterministic trend and a stochastic part) but if u_t is stationary, as assumed, then the latter is stationary in the TS model, so that the influence of past shocks dies out after one period, and nonstationary in the DS model.

Difference-stationary (DS) models

The second model is called "**difference-stationary**" as Δy_t is a stationary ARMA.

Alternatively, if a nonstationary time series has the property that if it is differentiated one or more times, the resulting series will be stationary, than this series is called **homogeneous**.

Difference-stationary (DS) models

If $w_t = y_t - y_{t\text{-}1} = \Delta y_t$ is stationary

than y_t is first-order homogeneous nonstationary.

If $w_t = \Delta^2 y_t = \Delta\, y_t - \Delta y_{t\text{-}1}$ is stationary

than y_t is second-order homogeneous nonstationary.

The first difference eliminates a linear trend, whereas a second difference can eliminate a quadratic trend.

Example:

Random walk process

$$y_t = y_{t-1} + \varepsilon_t$$

Differencing the Random walk process

$$w_t = \Delta y_t = y_t - y_{t-1} = \varepsilon_t$$

Since the ε_t are assumed to be independent over time, w_t is a stationary process (white noise).

So the random walk process is first-order homogeneous.

Under the assumption of a unit root, the DS model for this data sample is estimated in first differences as

$$(2)\quad \Delta Y_t = \underset{(0.001)}{0.003} + \underset{(0.074)}{0.369}\Delta Y_{t-1} + \hat{v}_t$$

$$\hat{\sigma}_v = 0.01035$$

This particular sample DS model will be denoted as the DS_{OLS} model.

Application: The Uncertain Unit Root in Real GNP

$$a(L)(1-L)y_t = \delta + b(L)\varepsilon_t$$

$$\Delta y_t = 0.003 + 0.369\Delta y_{t-1} + \varepsilon_t$$

$$(1-0.369L)\Delta y_t = 0.003 + \varepsilon_t$$

$$\underbrace{(1-0.369L)}_{a(L)}(1-L)y_t = 0.003 + \varepsilon_t$$

Alternatively : $y_t = 1369 y_{t-1} - 0.369 y_{t-2} + 0.003 + \varepsilon_t$

The estimated models (1) and (2) both appear to fit real GNP per capita fairly well; the standard deviations of their residuals are quite close, and plots of the residuals suggest no obvious outliers. In addition, Q statistics computed from the fitted residuals provide little evidence against the null hypothesis of no serial correlation at a variety of lags.

However, the estimated TS_{OLS} and DS_{OLS} models have very different implications for the persistence of the dynamic response of output to a random disturbance. To measure this persistence, consider the moving-average representation for the first difference of output implied by a TS or DS model:

$$(3) \quad \Delta Y_t = k + \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \dots$$

Recall: Wold Decomposition Theorem

Any stationary process can be written as

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-1} + \ldots = \mu + \psi(L)\varepsilon_t$$

$$\sum_{j=0}^{\infty} \left|\psi_j\right| < \infty, \text{roots of } \psi(z) = 0 \text{ are outside the unit circle}$$

ε_t~ white noise with mean zero and variance σ_ε^2

1. The unconditional expectation is a constant $E(y_t) = \mu$.
2. The forecast $\tilde{y}_{T+s|T} = E(y_{T+s}|y_T, y_{T-1}, \ldots)$ converges to the unconditional mean:

$$\lim_{s \to \infty} \tilde{y}_{T+s|T} = \mu$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Hence, we can write the stationary parts of the TS and DS models in this way

TS model

$$y_t - \delta_0 - \delta_1 t = \psi(L)\varepsilon_t$$

The mean is replaced by a linear function of date t.

DS model (Unit root process)

$$(1-L)y_t = \delta + \psi(L)\varepsilon_t \qquad \text{with } \psi(1) \neq 0$$

A prototypical example is obtained by setting $\psi(L) = 1$:

$$(1-L)y_t = \delta + \varepsilon_t \Leftrightarrow y_t = y_{t-1} + \delta + \varepsilon_t$$

→ **Random walk with drift**

Hamilton (1994) “Time Series Analysis”, p. 435-442

TS model

$$y_t - \delta_0 - \delta_1 t = \psi(L)\varepsilon_t$$

DS model (Unit root process)

$$(1-L)y_t = \delta + \psi(L)\varepsilon_t$$

We will be using this representation to derive differences between TS model and DS model in terms of

- impacts of past shocks on current levels of y_t
- forecast functions
- forecast error variances and forcast intervals

Hamilton (1994) “Time Series Analysis”, p. 435-442

TS model

$$y_t - \delta_0 - \delta_1 t = \psi(L)\varepsilon_t$$
$$= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-1} + \dots$$

s periods later..

$$y_{t+s} - \delta_0 - \delta_1 (t+s) = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \dots$$
$$+ \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \dots$$

Hence, $$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s$$

→ impacts (multipliers) are given by the Ψs.

Hamilton (1994) "Time Series Analysis", p. 435-442

for the **TS-Model**

$$y_t = -0.321 + 0.0003t + 1.335 y_{t-1} - 0.401 y_{t-2} + \varepsilon_t$$

$$(1 - 1.335L + 0.401L^2) y_t = -0.321 + 0.0003t + \varepsilon_t$$

$$\left(1 - 1.335L + 0.401L^2\right)\left(1 + \psi_1 L + \psi_2 L^2 + \ldots\right) = 1$$

$$1 - 1.335L + 0.401L^2 + \psi_1 L - 1.335\psi_1 L^2 + 0.401\psi_1 L^3$$
$$+ \psi_2 L^2 - 1.335\psi_2 L^3 + 0.401\psi_2 L^4 + \ldots = 1$$

$$-1.335 + \psi_1 = 0 \Rightarrow \psi_1 = 1.335$$

$$0.401 - 1.335\psi_1 + \psi_2 = 0$$

$$\Rightarrow \psi_2 = 1.335\psi_1 - 0.401 = 1.335^2 - 0.401 = 1.38$$

Similarly : $\psi_3, \psi_4, \ldots$

TS and DS model – impact of past shocks

TS model $y_t = \delta_0 + \delta_1 t + u_t \qquad u_t = \psi(L)\varepsilon_t$

$$y_t - \delta_0 - \delta_1 t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-1} + \ldots \qquad \rightarrow \qquad \frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s$$

In the **Example**: $y_t = -0.321 + 0.0003t + 1.335 y_{t-1} - 0.401 y_{t-2} + \varepsilon_t$

$$y_t = \underbrace{\left[(1-\varphi_1-\varphi_2)\delta_0 + (\varphi_1 + 2\varphi_2)\delta_1\right]}_{\delta} + \underbrace{\left[(1-\varphi_1-\varphi_2)\delta_1\right]}_{\gamma} t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

$$\Updownarrow$$

$y_t = \delta_0 + \delta_1 t + u_t \qquad$ with $\qquad u_t = \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \varepsilon_t$

$\rightarrow$ What is $\psi(L)$ for $u_t = 1.335 u_{t-1} - 0.401 u_{t-2} + \varepsilon_t$

Hamilton (1994) “Time Series Analysis”, p. 435-442

for the **TS-Model**

$$(1-1.335L+0.401L^2)u_t=\varepsilon_t$$

$$(1-1.335L+0.401L^2)(1+\psi_1 L+\psi_2 L^2+\dots)=1$$

$$1-1335L+0.401L^2+\psi_1 L-1335\psi_1 L^2+0.401\psi_1 L^3$$
$$+\psi_2 L^2-1335\psi_2 L^3+0.401\psi_2 L^4+\dots=1$$

$$-1.335+\psi_1=0\Rightarrow\psi_1=1.335$$

for the **TS-Model**

$$\psi_1 = 1.335$$

$$1 - 1335L + 0.401L^2 + \psi_1 L - 1335\psi_1 L^2 + 0.401\psi_1 L^3$$
$$+ \psi_2 L^2 - 1335\psi_2 L^3 + 0.401\psi_2 L^4 + \ldots = 1$$

$$0.401 - 1.335\psi_1 + \psi_2 = 0$$
$$\Rightarrow \psi_2 = 1.335\psi_1 - 0.401 = 1.335^2 - 0.401 = 1.38$$

for the **TS-Model** $\psi_1 = 1.335 \quad \psi_2 = 1.38$

$$\left(1 - 1.335L + 0.401L^2\right)\left(1 + \psi_1 L + \psi_2 L^2 + \ldots\right) = 1$$

$$\begin{aligned}
&1 - 1.335L + 0.401L^2 \\
&+ \psi_1 L - 1.335\psi_1 L^2 + 0.401\psi_1 L^3 \\
&+ \psi_2 L^2 - 1.335\psi_2 L^3 + 0.401\psi_2 L^4 \\
&+ \psi_3 L^3 - 1.335\psi_3 L^4 + 0.401\psi_3 L^5 \ldots = 1
\end{aligned}$$

$$\Rightarrow \psi_3 = 1.335\psi_2 - 0.401\psi_1 = 1.3086$$

$$\Rightarrow \psi_4 = 1.335\psi_3 - 0.401\psi_2 = 1.1931$$

$$\ldots$$

Application: The Uncertain Unit Root in Real GNP

TABLE 1—CUMULATIVE IMPULSE RESPONSES OF OLS MODELS

Model	Horizon (quarters) 1	2	4	8	12	16	20	30	40
DS_{OLS}	1.37 (0.07)	1.51 (0.13)	1.57 (0.17)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)
TS_{OLS}	1.33 (0.07)	1.38 (0.13)	1.19 (0.18)	0.73 (0.23)	0.43 (0.23)	0.25 (0.20)	0.15 (0.15)	0.04 (0.07)	0.01 (0.02)

Note: Standard errors are given in parentheses.

DS model

$$(1-L)y_t = \delta + \psi(L)\varepsilon_t$$

$$\Delta y_t = \delta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-1} + \ldots$$

s periods later..

$$\text{a}y_{t+s} = \delta + \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + \ldots$$
$$+ \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \ldots$$

Hence,

$$\frac{\partial\,\Delta y_{t+s}}{\partial\varepsilon_t} = \psi_s$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

for the **DS-Model** $\Delta y_t = 0.003 + 0.369\Delta y_{t-1} + \varepsilon_t$

$$\underbrace{(1-0.369L)}_{a(L)}(1-L)y_t = 0.003 + \varepsilon_t$$

$$(1-0.369L)(1+\psi_1 L + \psi_2 L^2 + \dots) = 1$$

$$1-0.369L + \psi_1 L - 0.369\psi_1 L^2 + \psi_2 L^2 - 0.369\psi_2 L^3 + \dots = 1$$

$$-0.369 + \psi_1 = 0 \Rightarrow \psi_1 = 0.369$$

$$-0.369\psi_1 + \psi_2 = 0 \Rightarrow \psi_2 = 0.369\psi_1 = 0.369^2 = 0.136$$

$$-0.369\psi_2 + \psi_3 = 0 \Rightarrow \psi_3 = 0.369\psi_2 = 0.369^3$$

Similarly : $\psi_4, \psi_5, \dots$

The Uncertain Unit Root in Real GNP

$$\text{(3)} \quad \Delta Y_t = k + \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \ldots$$

where k is some constant and ε_t is the innovation of the model. In this form, the sum of the a_i's measures the model response to a unit innovation.[5] A unit shock in period t affects ΔY_{t+h} by a_h and affects Y_{t+h} by $c_h \equiv 1 + a_1 + \cdots + a_h$. Thus, for various horizons, the cumulative response c_h answers the question: how does a shock today affect the level of real output in the short, medium, and long run? With quarterly data, for example, c_{20} measures the impact of a shock today on Y_t five years hence.

For the stationary part of the **DS model** we know

$$
\mathrm{a}y_{t+s} = \delta + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \ldots
$$
$$
+ \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \ldots
$$

and $$\frac{\partial\ \Delta y_{t+s}}{\partial \varepsilon_t} = \psi_s$$

How do we obtain the "cumulative response" $\dfrac{\partial\ y_{t+s}}{\partial \varepsilon_t}$?

We use the fact that

$$
\begin{aligned}
y_{t+s} &= \left(y_{t+s} - y_{t+s-1}\right) + \left(y_{t+s-1} - y_{t+s-2}\right) + \ldots + \left(y_{t+1} - y_t\right) + y_t \\
&= \Delta y_{t+s} + \Delta y_{t+s-1} + \ldots + \Delta y_{t+1} + y_t
\end{aligned}
$$

Hamilton (1994) "Time Series Analysis", p. 435-442

The level of the variable at date *t* + *s* is simply the sum of the changes between *t* and *s* + *t*:

$$\begin{aligned} y_{t+s} &= (y_{t+s} - y_{t+s-1}) + (y_{t+s-1} - y_{t+s-2}) + \ldots + (y_{t+1} - y_t) + y_t \\ &= \Delta y_{t+s} + \Delta y_{t+s-1} + \ldots + \Delta y_{t+1} + y_t \end{aligned}$$

Changes can be written as

$$\begin{aligned} \Delta y_{t+s} = \delta + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \ldots \\ + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \ldots \end{aligned}$$

or

$$\begin{aligned} \Delta y_{t+s-1} = \delta + \varepsilon_{t+s-1} + \psi_1 \varepsilon_{t+s-2} + \psi_2 \varepsilon_{t+s-3} + \ldots \\ + \psi_{s-1} \varepsilon_t + \psi_s \varepsilon_{t-1} + \ldots \end{aligned}$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

TS and DS model – impact of past shocks

Inserting the expressions for Δy_{t+s} , Δy_{t+s-1} ….. into

$$y_{t+s} = (y_{t+s} - y_{t+s-1}) + (y_{t+s-1} - y_{t+s-2}) + \ldots + (y_{t+1} - y_t) + y_t$$
$$= \Delta y_{t+s} + \Delta y_{t+s-1} + \ldots + \Delta y_{t+1} + y_t$$

and collecting terms (particularly for) gives

$$\begin{aligned} y_{t+s} &= \{\delta + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \ldots + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \ldots\} \\ &\quad + \{\delta + \varepsilon_{t+s-1} + \psi_1 \varepsilon_{t+s-2} + \psi_2 \varepsilon_{t+s-3} + \ldots + \psi_{s-1} \varepsilon_t + \psi_s \varepsilon_{t-1} + \psi_{s+1} \varepsilon_{t-2} + \ldots\} \\ &\quad + \ldots + \{\delta + \varepsilon_{t+1} + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \ldots\} + y_t \\ &= s\delta + y_t + \ldots + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\} \varepsilon_t \\ &\quad + \{\psi_{s+1} + \psi_s + \ldots + \psi_2\} \varepsilon_{t-1} + \ldots \end{aligned}$$

$$\Rightarrow \frac{\partial y_{t+s}}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial \varepsilon_t} + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\} = 1 + \psi_s + \psi_{s-1} + \ldots + \psi_1$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Application: The Uncertain Unit Root in Real GNP

sponse to a unit innovation.[5] A unit shock in period t affects ΔY_{t+h} by a_h and affects Y_{t+h} by $c_h \equiv 1 + a_1 + \cdots + a_h$. Thus, for

$$\Rightarrow \frac{\partial y_{t+s}}{\partial \varepsilon_t} = 1 + \psi_1 + \psi_2 + \ldots + \psi_s$$

TABLE 1—CUMULATIVE IMPULSE RESPONSES OF OLS MODELS

	Horizon (quarters)								
Model	1	2	4	8	12	16	20	30	40
DS_{OLS}	1.37 (0.07)	1.51 (0.13)	1.57 (0.17)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)
TS_{OLS}	1.33 (0.07)	1.38 (0.13)	1.19 (0.18)	0.73 (0.23)	0.43 (0.23)	0.25 (0.20)	0.15 (0.15)	0.04 (0.07)	0.01 (0.02)

Note: Standard errors are given in parentheses.

In the limit, the effect of a unit shock today on the level of output infinitely far in the future is given by c_∞. For any TS series, $c_\infty = 0$, because the effect of any shock is eliminated as reversion to the deterministic trend eventually dominates. For a DS series, $c_\infty \neq 0$; that is, each shock has some permanent effect. However, the impulse response of real output at an infinite horizon is of no practical economic significance; indeed, horizons of less than 10 years are usually of greatest interest. At these short

Application: The Uncertain Unit Root in Real GNP

usually of greatest interest. At these short horizons, the dynamic responses of TS and

DS models may be quite similar or quite different depending on the values taken by the parameters of the models. Thus, the presence of a unit root determines whether c_∞ is positive or zero, but it does not determine all of the model properties of economic interest. It is in this sense that, as

Application: The Uncertain Unit Root in Real GNP

TABLE 1—CUMULATIVE IMPULSE RESPONSES OF OLS MODELS

	Horizon (quarters)								
Model	1	2	4	8	12	16	20	30	40
DS_{OLS}	1.37 (0.07)	1.51 (0.13)	1.57 (0.17)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)	1.59 (0.19)
TS_{OLS}	1.33 (0.07)	1.38 (0.13)	1.19 (0.18)	0.73 (0.23)	0.43 (0.23)	0.25 (0.20)	0.15 (0.15)	0.04 (0.07)	0.01 (0.02)

Note: Standard errors are given in parentheses.

The Uncertain Unit Root in Real GNP

The estimated model responses are shown in Table 1, with standard errors in parentheses.[6] The impulse response of the DS_{OLS} model implies not only shock persistence but shock magnification. The effect of an innovation is not reversed through time, and it eventually increases the level of real GNP by more than one and a half times the size of the innovation ($c_{20} = 1.59$). In contrast, the TS model exhibits fairly rapid reversion to trend, with 85 percent of a shock dissipated after five years ($c_{20} = 0.15$). Thus, the cumulative impulse responses of these two models, each estimated from the same data sample, imply very different economic dynamics at cyclical frequencies. Because the TS_{OLS} and DS_{OLS} models of aggregate output have such different persistence properties, it would be useful to have a test capable of distinguishing between them. The next section explores the ability of one commonly used unit-root test to accomplish this task.

Comparison of Forecasts

TS model

The known deterministic component ($\alpha + \delta \cdot t$) is simply added to the forecast of the stationary stochastic component:

$$y_{T+s} = \delta_0 + \delta_1(T+s) + \varepsilon_{T+s} + \psi_1 \varepsilon_{T-s-1} + \psi_2 \varepsilon_{T+s-2} + \ldots$$
$$+ \psi_s \varepsilon_T + \psi_{s+1} \varepsilon_{T-1} + \ldots$$

$$\tilde{y}_{T+s|T} = E\left(y_{T+s} | y_T, \ldots, y_1\right)$$
$$= \delta_0 + \delta_1(T+s) + \psi_s \varepsilon_T + \psi_{s+1} \varepsilon_{T-1} + \ldots$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecasts

TS model

$$\begin{aligned}\tilde{y}_{T+s|T} &= E\left(y_{T+s}|y_T,\ldots,y_1\right)\\ &= \delta_0 + \delta_1(T+s) + \psi_s\varepsilon_T + \psi_{s+1}\varepsilon_{T-1} + \ldots\end{aligned}$$

As the forecast horizon (*s*) grows large, this forecast converges in mean square to the linear time trend.

$$E\left[\tilde{y}_{T+s|T} - \delta_0 - \delta_1(T+s)\right]^2 \to 0 \quad \text{as} \quad s \to \infty$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecasts TS and DS model

Comparison of Forecasts

DS model

From $$y_{T+s} = \Delta y_{T+s} + \Delta y_{T+s-1} + \ldots + \Delta y_{T+1} + y_T$$

and $$\Delta y_{T+s} = \delta + \varepsilon_{T+s} + \psi_1 \varepsilon_{T-s-1} + \psi_2 \varepsilon_{T+s-2} + \ldots$$
$$+ \psi_s \varepsilon_T + \psi_{s+1} \varepsilon_{T-1} + \ldots$$

it follows that

$$y_{T+s} = s\delta + y_T + \varepsilon_{T+s} + \ldots + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\}\varepsilon_T$$
$$+ \{\psi_{s+1} + \psi_s + \ldots + \psi_2\}\varepsilon_{T-1} + \ldots$$

Hamilton (1994) "Time Series Analysis", p. 435-442

Comparison of Forecasts

DS model (II)

$$y_{T+s} = s\delta + y_T + \varepsilon_{T+s} + \ldots + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\}\varepsilon_T + \{\psi_{s+1} + \psi_s + \ldots + \psi_2\}\varepsilon_{T-1} + \ldots$$

$$\begin{aligned}\tilde{y}_{T+s|T} &= E(y_{T+s}|y_T,\ldots,y_1) \\ &= s\delta + y_T + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\}\varepsilon_T \\ &\quad + \{\psi_{s+1} + \psi_s + \ldots + \psi_2\}\varepsilon_{T-1} + \ldots\end{aligned}$$

Reminder:
These are the ψs of Δy_t which is stationay

The forecast thus converges to a linear function of the horizon *s* with slope *δ*.

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecasts

TS versus DS model

The forecast $\hat{y}_{T+s|T}$ for both models converges to a linear function of the horizon s with slope δ.

The **key difference** is the **intercept** of the line. For a TS process, the forecast converges to a line whose intercept is the same regardless of the value of y_T.

By contrast, the intercept of the forecast for the DS process is continually changing with each new observation on y.

Hamilton (1994) “Time Series Analysis”, p. 435-442

TS and DS model

$$(1)\quad Y_t = \underset{(0.109)}{-0.321} + \underset{(0.00010)}{0.00030t} + \underset{(0.073)}{1.335Y_{t-1}}$$

$$\underset{(0.073)}{-0.401Y_{t-2}} + u_t \qquad \hat{\sigma}_u = 0.01013$$

TS model

$$y_t - \delta_0 - \delta_1 t = \psi(L)\varepsilon_t$$

DS model (Unit root process)

$$(1-L)y_t = \delta + \psi(L)\varepsilon_t$$

$$(2)\quad \Delta Y_t = \underset{(0.001)}{0.003} + \underset{(0.074)}{0.369\Delta Y_{t-1}} + \hat{v}_t$$

$$\hat{\sigma}_v = 0.01035$$

We will be using this re
between TS model and

- impacts of past shocks on current levels of y_t
- forecast functions
- forecast error variances and forcast intervals

Hamilton (1994) “Time Series Analysis”, p. 435-442

Prediction Intervals for TS and DS models

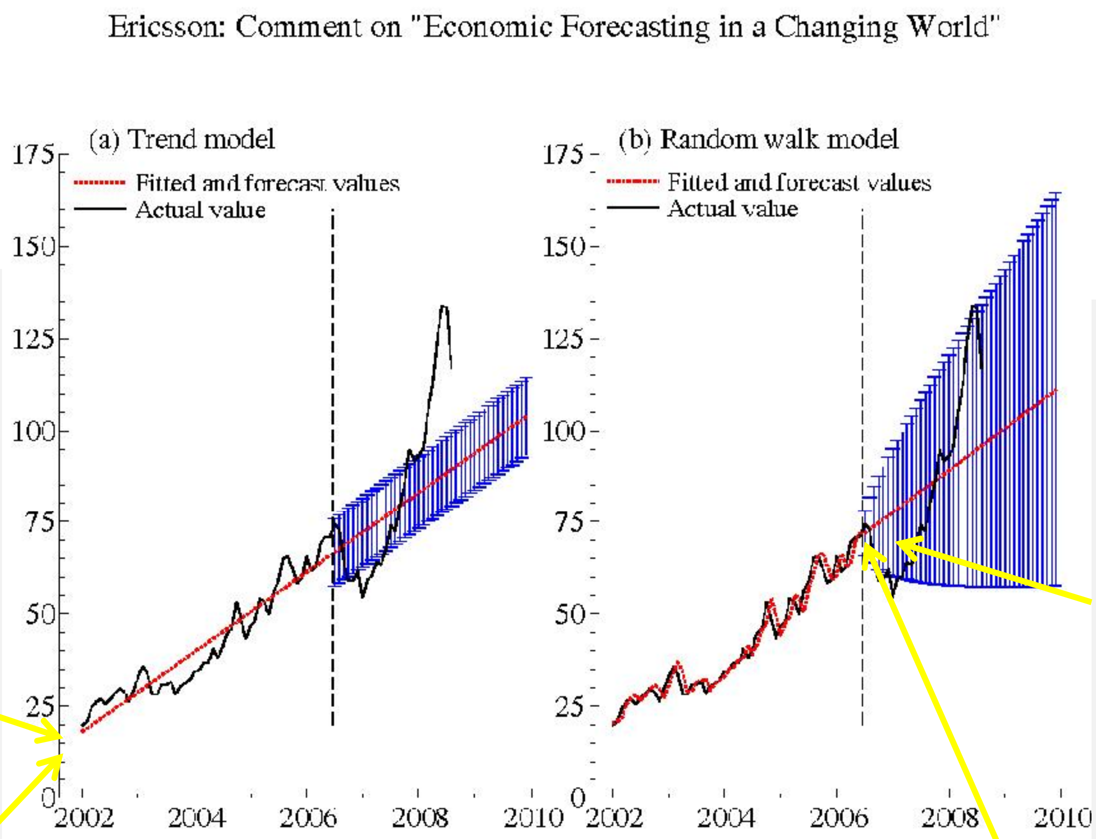


Figure 2: Actual, fitted, and forecast values from the trend and random walk models of the oil price, with 95% confidence intervals for the forecasts.

The **key difference** is the **intercept** of the line.

For a TS process, the forecast converges to a line whose intercept is the same regardless of the value of y_T.

The **key difference** is the **intercept** of the line.

By contrast, the intercept of the forecast for the DS process is continually changing with each new observation on y.

$$\tilde{y}_{T+s|T} = \delta_0 + \delta_1(T+s) + \psi_s \varepsilon_T + \psi_{s+1}\varepsilon_{T-1} + \ldots$$

$$\tilde{y}_{T+s|T} = s\delta + y_T + \{\psi_s + \psi_{s-1} + \ldots + \psi_1\}\varepsilon_T + \{\psi_{s+1} + \psi_s + \ldots + \psi_2\}\varepsilon_{T-1} + \ldots$$

In the random walk case all ψs are 0.

Comparison of Forecast Errors

TS model (I)

The *s*-period-ahead forecast error is:

$$
\begin{aligned}
y_{T+s|T} - \hat{y}_{T+s|T} &= \left\{\delta_0 + \delta(T+s) + \varepsilon_{T+s} + \psi_1\varepsilon_{T+s-1} + \psi_2\varepsilon_{T+s-2} + \dots \right. \\
&\quad \left. + \psi_{s-1}\varepsilon_{T+1} + \psi_s\varepsilon_T + \psi_{s+1}\varepsilon_{T-1} + \dots \right\} \\
&\quad - \left\{\delta_0 + \delta(T+s) + \psi_s\varepsilon_T + \psi_{s+1}\varepsilon_{T-1} + \psi_{s+2}\varepsilon_{T-2} + \dots \right\} \\
&= \varepsilon_{T+s} + \psi_1\varepsilon_{T+s-1} + \psi_2\varepsilon_{T+s-2} + \dots + \psi_{s-1}\varepsilon_{T+1}
\end{aligned}
$$

MSE of this forecast:

$$
E\left(y_{T+s|T} - \hat{y}_{T+s|T}\right)^2 = \left\{1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2\right\}\sigma_\varepsilon^2
$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecast Errors

TS model (II)

The *MSE* increases with the horizon *s*, though as *s* becomes large, the added uncertainty from forecasting farther into the future becomes negligible:

$$\lim_{s\to\infty} E\left(y_{t+s|t} - \hat{y}_{t+s|t}\right)^2 = \left\{1 + \psi_1^2 + \psi_2^2 + \ldots\right\}\sigma_\varepsilon^2$$

Note that the limiting *MSE* is just the unconditional variance of the stationary component $\psi(L)\varepsilon_t$.

Hamilton (1994) "Time Series Analysis", p. 435-442

Comparison of Forecast Errors

DS model (I)

The *s*-period-ahead forecast error is:

$$
\begin{aligned}
y_{T+s|T} - \hat{y}_{T+s|T} &= \left\{\Delta y_{T+s} + \Delta y_{T+s-1} + \ldots + \Delta y_{T+1} + y_T\right\} \\
&\quad - \left\{\Delta \hat{y}_{T+s|T} + \Delta \hat{y}_{T+s-1|T} + \ldots + \Delta \hat{y}_{T+1|t} + y_T\right\} \\
&= \left\{\Delta y_{T+s} - \Delta \hat{y}_{T+s|T}\right\} + \left\{\Delta y_{T+s-1} - \Delta \hat{y}_{T+s-1|T}\right\} + \ldots + \left\{\Delta y_{T+1} - \Delta \hat{y}_{T+1|T}\right\}
\end{aligned}
$$

$$\Delta y_{T+s} = \delta + \varepsilon_{T+s} + \psi_1 \varepsilon_{T-s-1} + \psi_2 \varepsilon_{T+s-2} + \ldots + \psi_s \varepsilon_T + \psi_{s+1} \varepsilon_{T-1} + \ldots$$

$$\Delta \hat{y}_{T+s|T} = E\left[\Delta y_{T+s} \mid \Omega_T\right] = \delta + \psi_s \varepsilon_T + \psi_{s+1} \varepsilon_{T-1} + \ldots$$

$$\Delta y_{T+s} - \Delta \hat{y}_{T+s|T} = \varepsilon_{T+s} + \psi_1 \varepsilon_{T-s-1} + \psi_2 \varepsilon_{T+s-2} + \ldots + \psi_{s-1} \varepsilon_{T+1}$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecast Errors TS and DS model

Comparison of Forecast Errors

DS model (I)

The *s*-period-ahead forecast error is:

$$y_{T+s|T} - \hat{y}_{T+s|T} = \left\{\Delta y_{T+s} - \Delta\hat{y}_{T+s|T}\right\} + \left\{\Delta y_{T+s-1} - \Delta\hat{y}_{T+s-1|T}\right\} + \ldots + \left\{\Delta y_{T+1} - \Delta\hat{y}_{T+1|T}\right\}$$

$$\begin{aligned}
\Delta y_{T+s} - \Delta\hat{y}_{T+s|T} &= \varepsilon_{T+s} + \psi_1\varepsilon_{T+s-1} + \psi_2\varepsilon_{T+s-2} + \ldots + \psi_{s-2}\varepsilon_{T+2} + \psi_{s-1}\varepsilon_{T+1} \\
\Delta y_{T+s-1} - \Delta\hat{y}_{T+s-1|T} &= \varepsilon_{T+s-1} + \psi_1\varepsilon_{T-s-2} + \psi_2\varepsilon_{T+s-3} + \ldots + \psi_{s-2}\varepsilon_{T+1} \\
\vdots \qquad &\ \vdots \qquad \ddots \\
\Delta y_{T+1} - \Delta\hat{y}_{T+1|T} &= \varepsilon_{T+1}
\end{aligned}$$

$$y_{T+s|T} - \hat{y}_{T+s|T} = \varepsilon_{T+s} + \boxed{\{1+\psi_1\}\varepsilon_{T+s-1}} + \{1+\psi_1+\psi_2\}\varepsilon_{T+s-2} + \ldots + \boxed{\{1+\psi_1+\psi_2+\ldots+\psi_{s-1}\}\varepsilon_{T+1}}$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecast Errors

DS model (I)

The *s*-period-ahead forecast error is:

$$y_{T+s|T} - \hat{y}_{T+s|T} = \varepsilon_{T+s} + \{1+\psi_1\}\varepsilon_{T+s-1} + \{1+\psi_1+\psi_2\}\varepsilon_{T+s-2} + \ldots + \{1+\psi_1+\psi_2+\ldots+\psi_{s-1}\}\varepsilon_{T+1}$$

$$\begin{aligned}\left(y_{T+s|T} - \hat{y}_{T+s|T}\right)^2 = &\left(\varepsilon_{T+s} + \{1+\psi_1\}\varepsilon_{T+s-1} + \{1+\psi_1+\psi_2\}\varepsilon_{T+s-2} + \ldots + \{1+\psi_1+\psi_2+\ldots+\psi_{s-1}\}\varepsilon_{T+1}\right)\\ &\cdot\left(\varepsilon_{T+s} + \{1+\psi_1\}\varepsilon_{T+s-1} + \{1+\psi_1+\psi_2\}\varepsilon_{T+s-2} + \ldots + \{1+\psi_1+\psi_2+\ldots+\psi_{s-1}\}\varepsilon_{T+1}\right)\end{aligned}$$

$$E\left[\left(y_{T+s|T} - \hat{y}_{T+s|T}\right)^2 \mid \Omega_T\right] = \left\{1 + \left(1+\psi_1\right)^2 + \left(1+\psi_1+\psi_2\right)^2 + \ldots \; + \left(1+\psi_1+\psi_2+\ldots+\psi_{s-1}\right)^2\right\}\sigma_\varepsilon^2$$

Hamilton (1994) “Time Series Analysis”, p. 435-442

Comparison of Forecast Errors

DS model (II)

MSE of this forecast:

$$E\left(y_{T+s|T}-\hat{y}_{T+s|T}\right)^2=\left\{1+\left(1+\psi_1\right)^2+\left(1+\psi_1+\psi_2\right)^2+\ldots\right.$$
$$\left.+\left(1+\psi_1+\psi_2+\ldots+\psi_{s-1}\right)^2\right\}\sigma_\varepsilon^2$$

The *MSE* increases with the length of the forecasting horizon *s*, though in contrast to the trend-stationary case, the *MSE* does not converge to any fixed value as *s* goes to infinity. Instead, it asymptotically approaches a linear function of s with slope $\left(1+\psi_1+\psi_2+\ldots\right)^2\sigma_\varepsilon^2$.

Hamilton (1994) “Time Series Analysis”, p. 435-442

Prediction Intervals for TS model

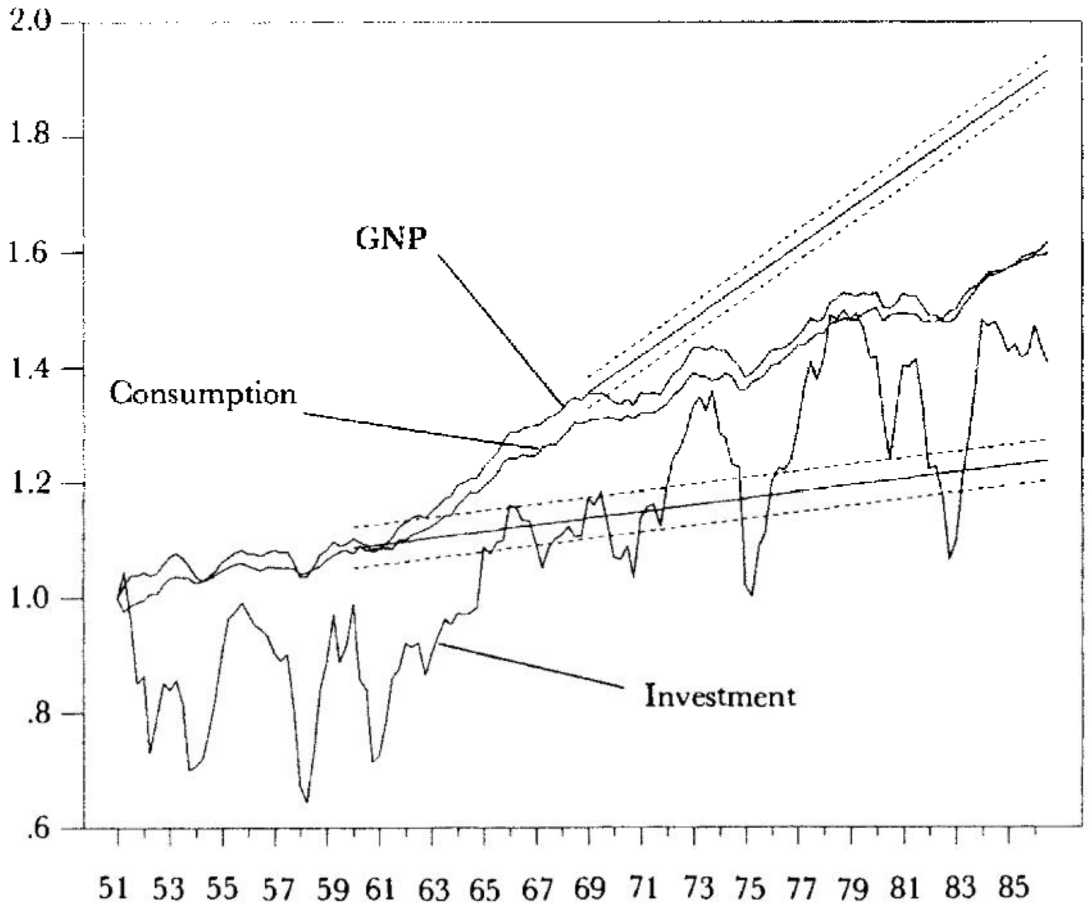


Fig. 1. Postwar real per capita U.S. GNP, total consumption, and gross private domestic investment (in logarithms)

Prediction Intervals for TS and DS models

Ericsson: Comment on "Economic Forecasting in a Changing World"

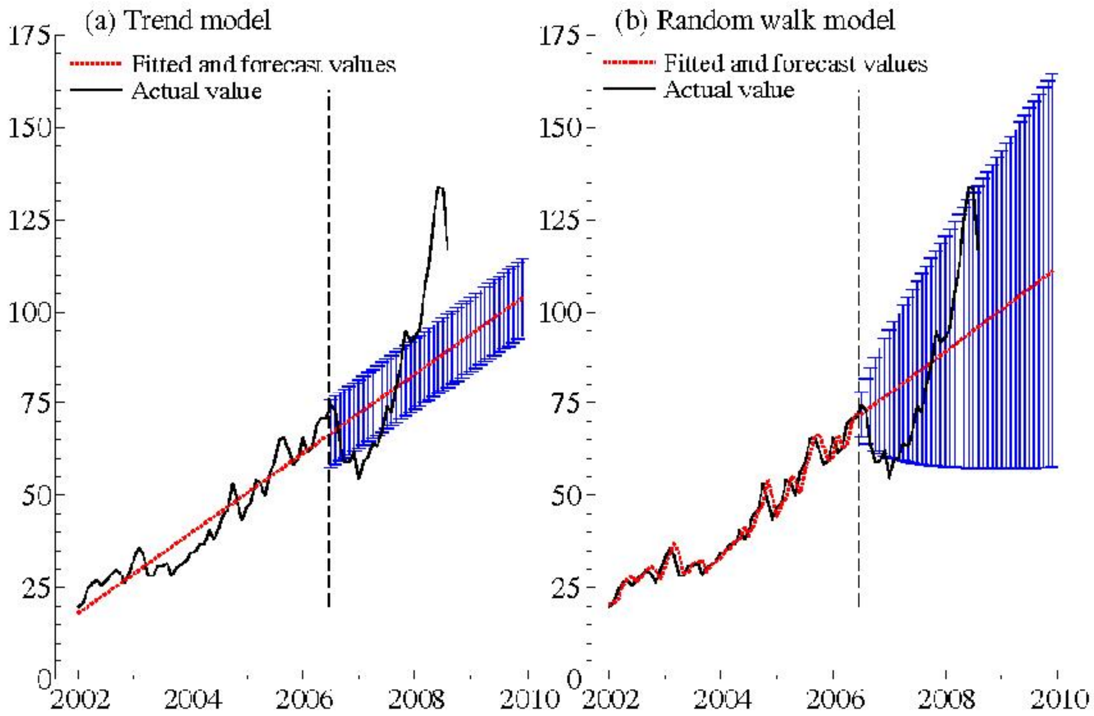


Figure 2: Actual, fitted, and forecast values from the trend and random walk models of the oil price, with 95% confidence intervals for the forecasts.

TS and DS model

The **TS model** and the **DS model** are both nonstationary and both may have a linear deterministic trend.

However, we have seen that they differ in terms of

- impacts of past shocks on current levels of y_t
- forecast functions
- forecast error variances and forcast intervals

They further differ in terms of

- the appropriate way to remove the trend

Hamilton (1994) “Time Series Analysis”, p. 435-442

The **trend-stationary model** can also be made stationary by **first differencing**:

$$y_t = \delta_0 + \delta_1 t + u_t$$

$$y_t - y_{t-1} = \delta_1 [t - (t-1)] + u_t - u_{t-1}$$

$$\Delta y_t = \delta_1 + \Delta u_t$$

However, it induces “new”, artificial serial correlation into the error term of the differenced model.

Example: u_t ~ white noise

$$\Delta u_t = u_t - u_{t-1}$$

is an MA(1) with $\theta_1 = 1$

The opposite, however, is not true: **removing the deterministic trend** from **an ARIMA model with a constant term** does not make the series stationary.

Example: Random walk with drift

$$y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_t \Rightarrow \underbrace{y_t - \delta t}_{x_t} = y_0 + \sum_{j=1}^{t} \varepsilon_t$$

For a fixed value y_0 the series x_t has a constant mean:

$$E(x_t) = y_0$$

but the variance depends on t: $Var(x_t) = t \cdot \sigma_\varepsilon^2$

Linear trend is always significant..

Data is generated as:.. $y_t = 5 + 1 \cdot t + \varepsilon_t$ $\quad y_t = y_{t-1} + \varepsilon_t$ $\quad y_t = y_{t-1} + \delta + \varepsilon_t$

Tabelle 5.1: ***Ergebnisse linearer Trendbereinigung***
(100 Beobachtungen)

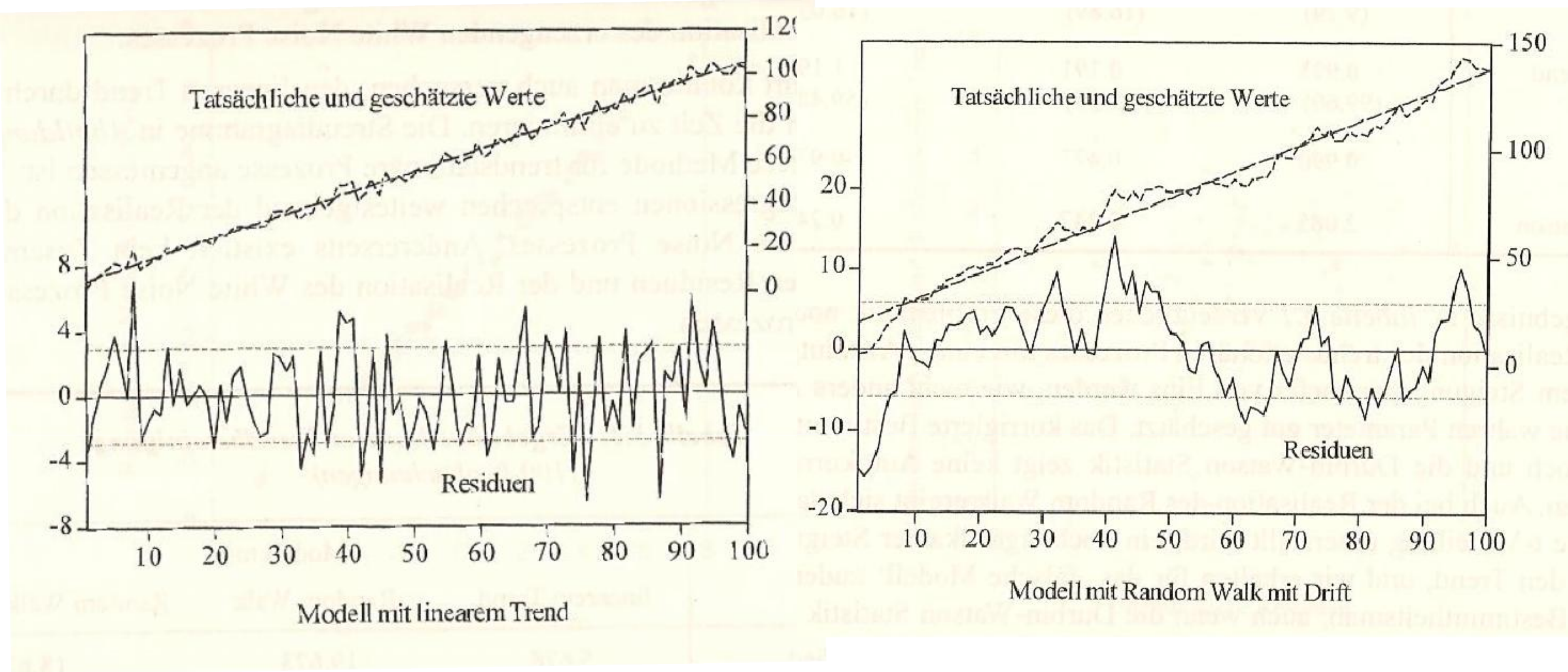
$y_t = \hat{\delta}_0 + \hat{\delta}_1 t + \hat{u}_t$	Modell mit linearem Trend	Modell mit Random Walk	Modell mit Random Walk mit Drift
Absolutglied $\hat{\delta}_0$	5.678 (9.79)	19.673 (16.89)	18.673 (16.03)
linearer Trend $\hat{\delta}_1$	0.993 (99.60)	0.191 (9.55)	1.191 (59.48)
$\bar{R}^2$	0.990	0.477	0.973
Durbin-Watson	2.085	0.247	0.247

$$t = \frac{\hat{¢}_1}{\hat{\sigma}_{\hat{¢}_1}}$$

$DW \approx 2(1 - \hat{\rho}_1)$ where $\hat{\rho}_1$ is the 1st order autocorrelation of $\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\delta}_1 t$

If $\hat{\rho}_1 \approx 0$ then $DW \approx 2$. If $\hat{\rho}_1 \approx 1$ then $DW \approx 0$.

… but residual behavior is different...



„Man sieht, dass die Residuen beim Random Walk mit Drift noch deutlich systematische Bewegungen enthalten, die fälschlicherweise als genuine Zyklen interpretiert werden könnten.“

Kirchgässner & Wolters (2006) “Einführung..”, p. 145

TS and DS model

We have seen that the **TS model** and the **DS model** are both nonstationary and both may have a linear deterministic trend but differ in terms of

- impacts of past shocks on current levels of y_t
- forecast functions
- forecast error variances and forcast intervals
- the appropriate way to remove the trend

Hence, we want to distinguish between them

→ unit root test (see below)

However, some nonstationary processes are neither TS nor DS.

Hamilton (1994) “Time Series Analysis”, p. 435-442

There are nonstationary models and time series that are **neither TS nor DS models.**

Example: AR(1)-Process with $\varphi_1 > 1$

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t$$

Repeated substitution yields:

$$y_t = \varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \varepsilon_j$$

If $\varphi_1 > 1$ then the series will show an explosive behavior, even if $y_0 = 0$.

Neither TS nor DS

Example: AR(1)-Process $\left(y_t = \varphi_1 y_{t-1} + \varepsilon_t\right)$ with $\varphi_1 > 1$

$$y_t = \varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \varepsilon_j$$

$$E(y_t) = E\left(\varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \varepsilon_j \right) = E\left(\varphi_1^t \cdot y_0\right) + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot E\left(\varepsilon_j\right)$$

$$= \varphi_1^t \cdot y_0 = \mu_t$$

The mean is not independent of time. The mean is exponentially increasing with time.

Neither TS nor DS

Example: AR(1)-Process $\left(y_t = \varphi_1 y_{t-1} + \varepsilon_t\right)$ with $\varphi_1 > 1$

$$y_t = \varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

$$Var\left(y_t\right) = Var\left(\varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j\right) = \sum_{j=1}^{t} \varphi_1^{2\cdot(t-j)} \cdot Var\left(\varepsilon_j\right)$$

$$= \left(\varphi_1^{2\cdot(t-1)} + \varphi_1^{2\cdot(t-2)} + \ldots + \varphi_1^4 + \varphi_1^2 + 1\right) \cdot \sigma_\varepsilon^2$$

$$= \frac{\varphi_1^{2t} - 1}{\varphi_1^2 - 1} \cdot \sigma_\varepsilon^2$$

The variance is increasing with time.

ARIMA model with a unit root and no constant term

$$a(L)(1-L)y_t = b(L)\varepsilon_t$$

Most “famous” special case:

Random walk without drift $\qquad y_t = y_{t-1} + \varepsilon_t$

Repeated substitution yields:

$$y_t = y_0 + \underbrace{\sum_{j=1}^{t} \varepsilon_j}_{\text{"stochastic trend"}}$$

No “built-in determinstic linear trend” but stochastic trend due to accumulated past shocks

ARIMA model with a unit root and no constant term

$$a(L)(1-L)y_t = b(L)\varepsilon_t$$

Another example: ARIMA(0,1,1)

$$y_t = y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1} \qquad \Delta y_t = \varepsilon_t - \theta_1\varepsilon_{t-1}$$

Since $y_t = \sum_{i=-\infty}^{t} \Delta y_i = \sum_{i=-\infty}^{0} \Delta y_i + \sum_{i=1}^{t} \Delta y_i = y_0 + \Delta y_1 + \Delta y_2 + \ldots + \Delta y_t$

$$y_t = y_0 + \underbrace{(1-\theta_1)\sum_{j=1}^{t-1}\varepsilon_j}_{\text{"stochastic trend"}} + \varepsilon_t + \theta_1\varepsilon_0$$

➔ ARIMA(0,1,1) also has “stochastic trend”

Unit root processes (DS models) with no constant term

ARIMA model with a unit root and no constant term

$$a(L)(1-L)y_t = b(L)\varepsilon_t$$

Any unit root process with no constant term has such a stochastic trend component (accumulated shocks).

This term was shown to be responsible for widening of forecast intervals in the case with constant term.

➔ Processes with a unit root but no constant term share this feature.

Hence, we may want to discriminate between them and their stationary counterpart (stationary ARMA model with no deterministic trend)

Stationary ARMA or ARIMA without constant term?

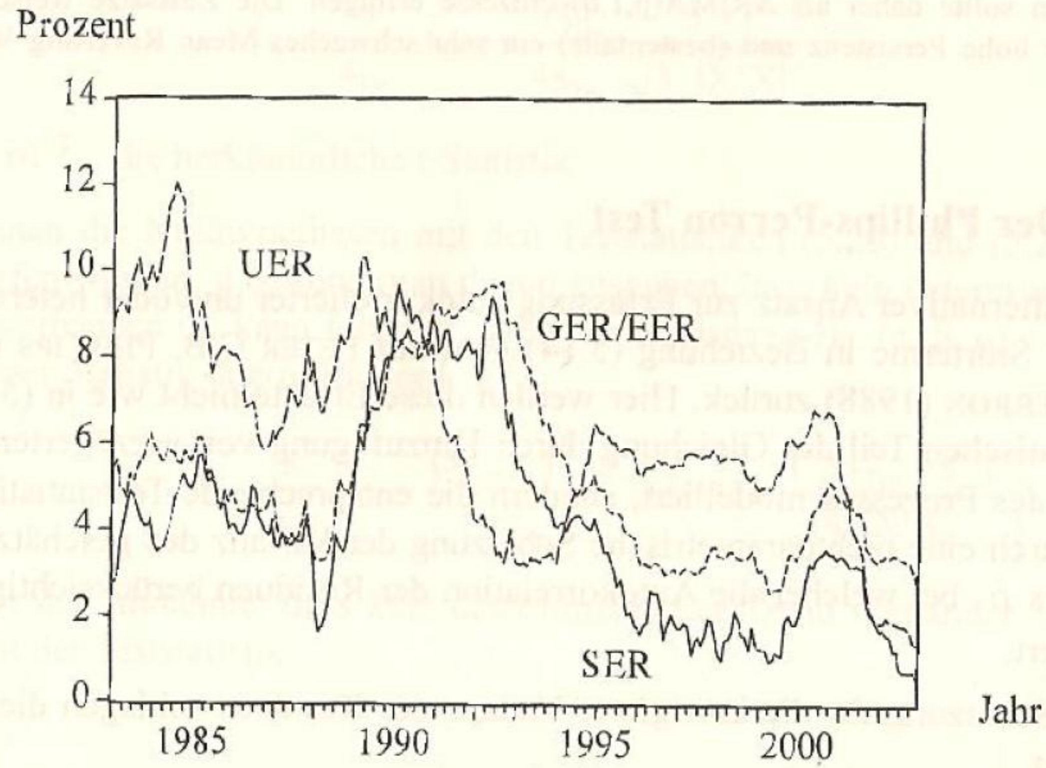


Abbildung 5.8: Entwicklung der schweizerischen, deutschen/europäischen und amerikanischen Euromarktsätze, Monatsdaten, Januar 1983 – Dezember 2002

Stationarity and the Autocorrelation Function

How to decide whether a time series is stationary or determine the appropriate number of times a homogenous nonstationary series should be differentiated to arrive at a stationary series?

The autocorrelation function for a stationary series drops off as the number of lags becomes large.

This is not the case for a nonstationary series.

Example AR(1): $y_t = \varphi_1 y_{t-1} + \varepsilon_t$

Variance and covariance for a fixed y_0 and φ = 1:

$$Var(y_t) = \gamma_0 = t \cdot \sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = \gamma_k = (t-k) \cdot \sigma_\varepsilon^2$$

Autocorrelation function:

$$\rho_\tau = \frac{(t-k)}{t} \approx 1$$ for small numbers of k

Johnston, DiNardo (1997) "Econometric Methods"

Identification in Practice

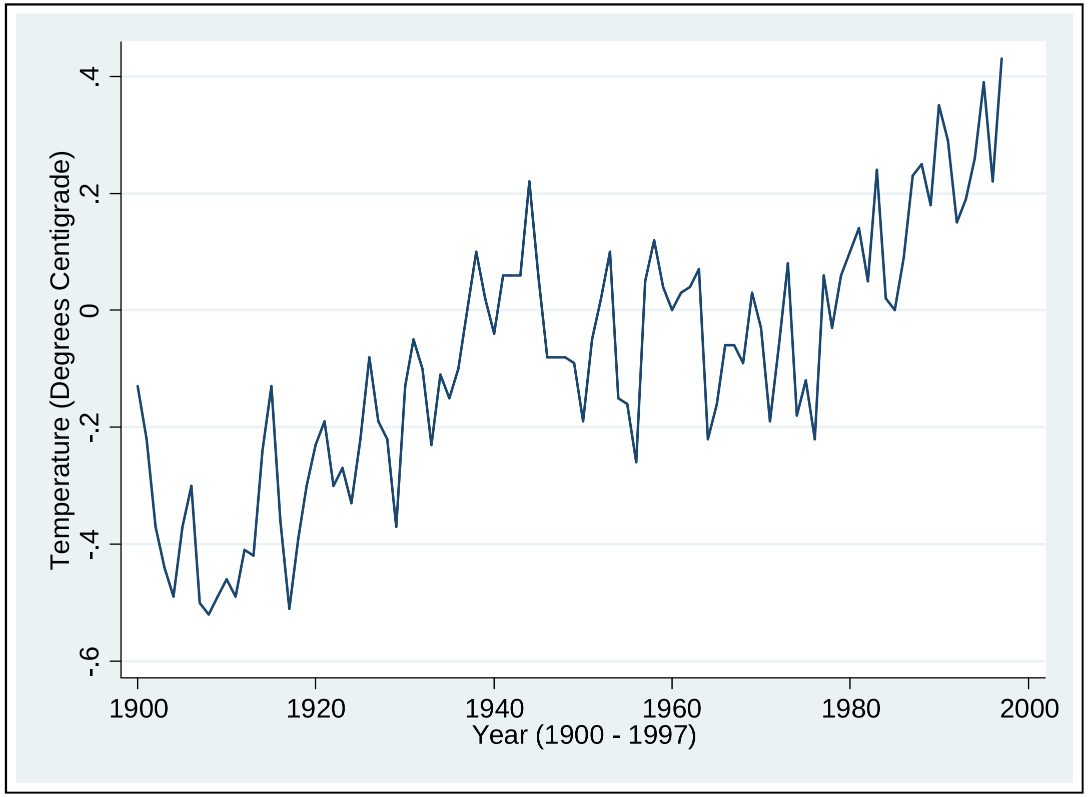
A time plot of the data will typically suggest whether any differencing is needed. If differencing is called for, then difference once and then inspect the time plot of the first difference.

Shumway/Stoffer (2000) “Time Series Analysis and Its Applications”, p. 145

Failure of the sample ACF and PACF to die out quickly at high lags and the appearance of smooth behavior in these quantities at high lags is an indication that further differencing is required.

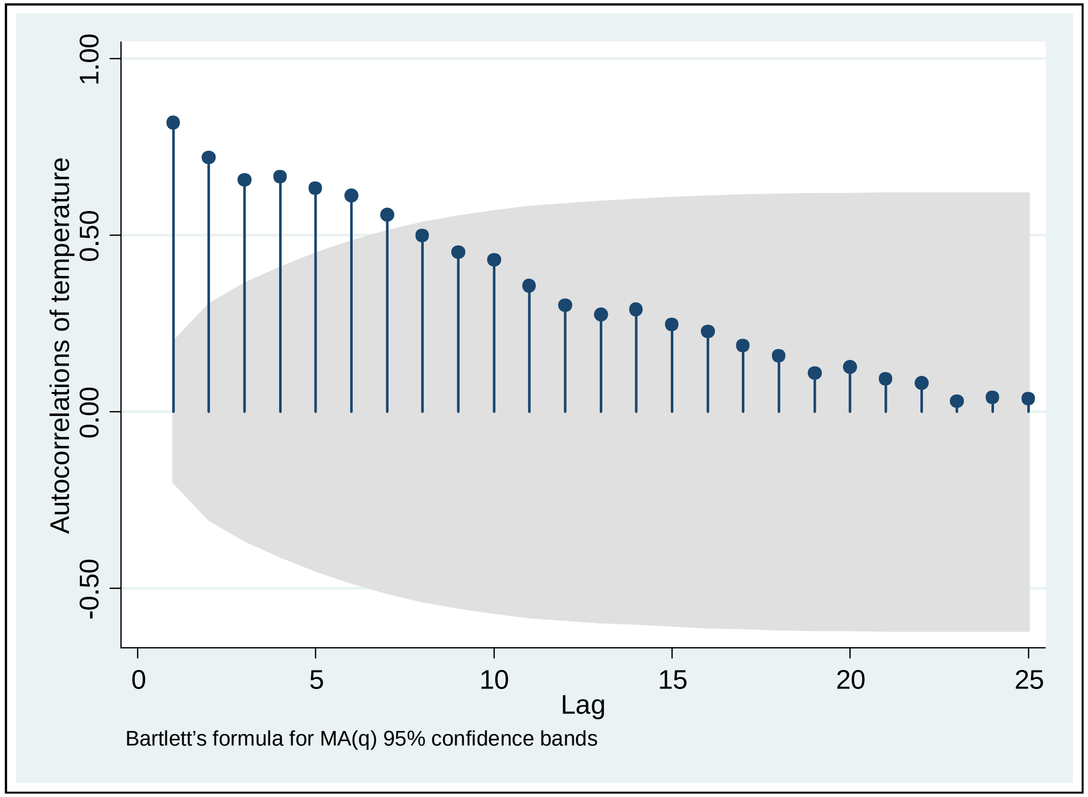
Granger/Newbold (1986) “Forecasting Economic Time Series”, p. 81

Example: Global warming data



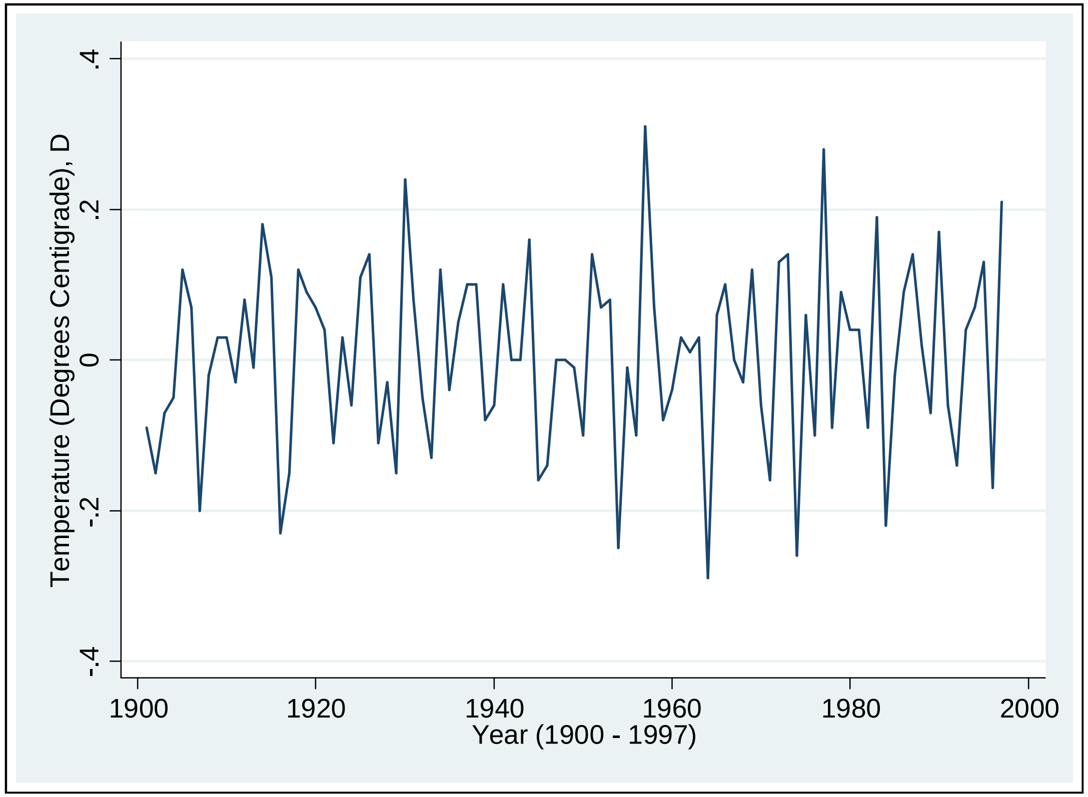
Shumway/Stoffer (2000) "Time Series Analysis and its Applications"

Autocorrelation Function of Global warming data



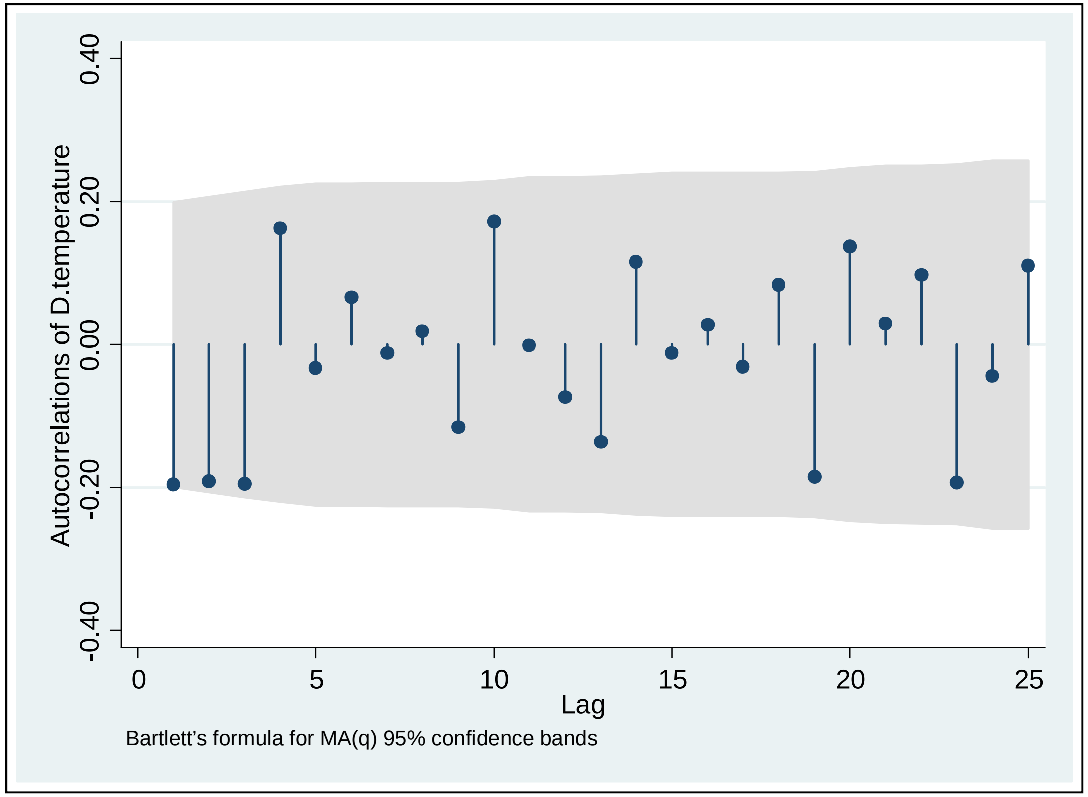
Shumway/Stoffer (2000) "Time Series Analysis and its Applications"

Differenced Global warming data



Shumway/Stoffer (2000) “Time Series Analysis and its Applications”

Autocorrelation Function of the differenced Global warming data



Shumway/Stoffer (2000) "Time Series Analysis and its Applications"

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • **Unit Root Tests** • Seasonal ARIMA

Principal Methods for **Detecting Nonstationarity**

1. Subjective judgment applied to the time series graph of the series and its correlogram
2. Formal statistical test for unit roots

Johnston, DiNardo (1997) “Econometric Methods, p. 215”

Testing for Unit Roots for a simple AR(1):

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

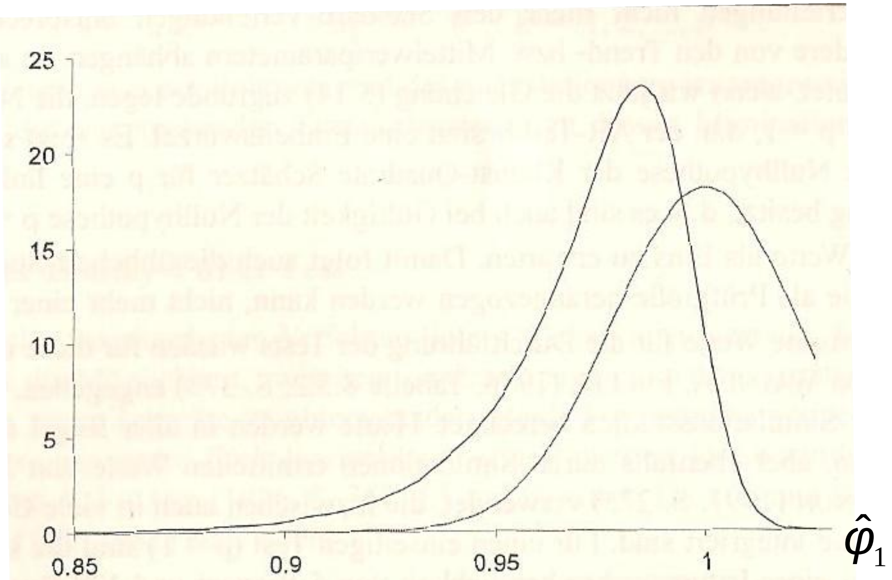
- a test for a unit root is a test for $\varphi_1 = 1$ (H_0)
- to test H_0, use OLS estimate $\hat{\varphi}_1$ and its standard error
- but, under H_0, standard *t*-ratio is not *t*-distributed
- the distribution is skewed to the right

  → the critical values are smaller than N(0,1) or *t* distr.
- using standard critical values will reject H_0 (unit root) to often
- **so use Dickey-Fuller *t*-tests**

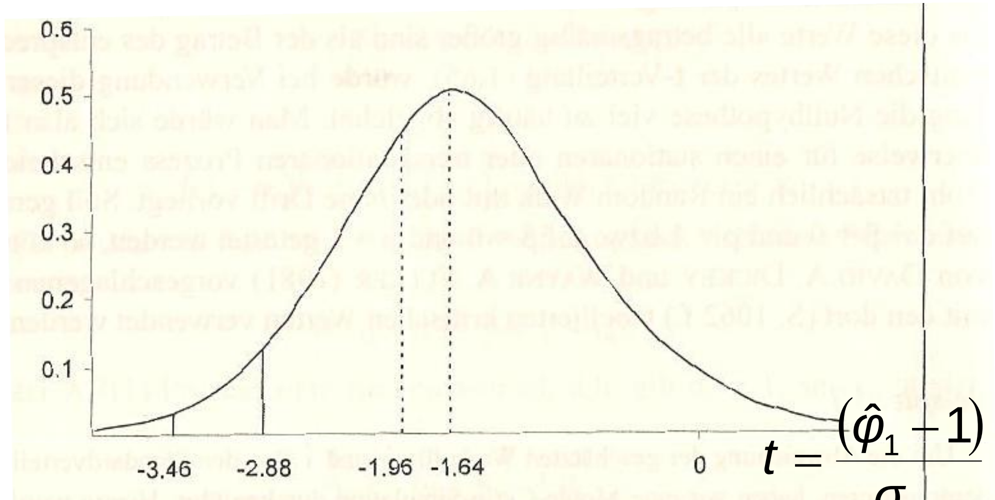
Verbeek (2000) "A Guide to Modern Econometrics "

Distribution of OLS slope and t under a unit root

Data is generated according to a simple random walk. T=200, 100000 replications



Dichte der geschätzten Koeffizienten im Vergleich mit einer Normalverteilung mit gleicher Varianz und $\mu = 1$



Dichte der Dickey-Fuller t-Statistik

$$t = \frac{(\hat{\varphi}_1 - 1)}{\sigma_{\hat{\varphi}_1}}$$

ng 5.7: Dichten des geschätzten Autokorrelationskoeffizienten und der t-Statistik unter der Nullhypothese eines Random Walk

Testing for Unit Roots

no constant, no trend	constant, no trend	constant and trend
$y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$
$H_0{:}\varphi_1 = 1$ $y_t = y_{t-1} + \varepsilon_t$ $H_1{:}\lvert\varphi_1\rvert < 1$ $y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$H_0{:}\varphi_1 = 1, (\delta = 0)$ $y_t = y_{t-1} + \varepsilon_t$ $H_1{:}\lvert\varphi_1\rvert < 1, (\delta \neq 0)$ $y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$	$H_0{:}\varphi_1 = 1, (\gamma = 0)$ $y_t = y_{t-1} + \delta + \varepsilon_t$ $H_1{:}\lvert\varphi_1\rvert < 1, (\gamma \neq 0)$ $y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$
• H_0: pure random walk (no drift) • H_1: stationary AR(1) with mean zero (i.e. strictly speaking $0 \leq \varphi_1 < 1$) • simplest case, mostly educational value • “Testing with zero intercept is extremely restrictive, so much that it is hard to imagine ever using it with economic time series”*	• H_0: pure random walk (no drift) • H_1: stationary AR(1) with arbitrary mean • applies to non-growing series • typical examples: “rates” (interest rates, inflation rates, unemployment rates)	• H_0: random walk with drift • H_1: trend stationary model with AR(1) errors • applies to growing series (but not explosive) • typical examples: GDP, consumption, investment

* Davidson, MacKinnon (1993) “Estimation and inference in econometrics ”, p.702

no constant, no trend	constant, no trend	constant and trend
$y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$
$H_0: \varphi_1 = 1$	$H_0: \varphi_1 = 1, (\delta = 0)$	$H_0: \varphi_1 = 1, (\gamma = 0)$
$y_t = y_{t-1} + \varepsilon_t$	$y_t = y_{t-1} + \varepsilon_t$	$y_t = y_{t-1} + \delta + \varepsilon_t$
$H_1: \varphi_1 < 1$	$H_1: \varphi_1 < 1, (\delta \neq 0)$	$H_1: \varphi_1 < 1, (\gamma \neq 0)$
$y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$
Estimating equations	Estimating equations	Estimating equations
$y_t = \varphi_1 y_{t-1} + \varepsilon_t$ or	$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$ or	$y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$ or
$\Delta y_t = \theta y_{t-1} + \varepsilon_t$	$\Delta y_t = \theta y_{t-1} + \delta + \varepsilon_t$	$\Delta y_t = \theta y_{t-1} + \delta + \gamma t + \varepsilon_t$
$\theta = (\varphi_1 - 1)$	$\theta = (\varphi_1 - 1)$	$\theta = (\varphi_1 - 1)$
Test statistics	Test statistics	Test statistics
$t = \frac{(\hat{\varphi}_1 - 1)}{\sigma_{\hat{\varphi}_1}}$ or $t = \frac{\hat{\theta}}{\sigma_{\hat{\theta}}}$	$t = \frac{(\hat{\varphi}_1 - 1)}{\sigma_{\hat{\varphi}_1}}$ or $t = \frac{\hat{\theta}}{\sigma_{\hat{\theta}}}$	$t = \frac{(\hat{\varphi}_1 - 1)}{\sigma_{\hat{\varphi}_1}}$ or $t = \frac{\hat{\theta}}{\sigma_{\hat{\theta}}}$

* Davidson, MacKinnon (1993) "Estimation and inference in econometrics ", p.702

Critical values for Dickey-Fuller tests

Sample Size T	No constant, no trend		Constant, no trend		Constant, trend	
	1%	5%	1%	5%	1%	5%
25	-2.66	-1.95	-3.75	-3.00	-4.38	-3.60
50	-2.62	-1.95	-3.58	-2.93	-4.15	-3.50
100	-2.60	-1.95	-3.51	-2.89	-4.04	-3.45
250	-2.58	-1.95	-3.46	-2.88	-3.99	-3.43
500	-2.58	-1.95	-3.44	-2.87	-3.98	-3.42
∞	-2.58	-1.95	-3.43	-2.86	-3.96	-3.41

Verbeek (2000) “A Guide to Modern Econometrics ”

Why one-sided tests? (H_1: $\varphi_1 < 1$)

If $\varphi_1 > 1$ in $y_t = \varphi_1 y_{t-1} + \varepsilon_t$

then repeated substitution implies

$$y_t = \varphi_1^t \cdot y_0 + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

and therefore $E[y_t] = \varphi_1^t \cdot y_0$, i.e. exponential growth of the mean. This is uncommon (or undone by taking logs) and therefore ruled out a-priori by standard DF-tests.

Why is the alternative in the "constant and trend"-case a TS-model with AR(1) errors?

TS-model $y_t = \delta_0 + \delta_1 t + u_t$ with AR(1) errors $u_t = \varphi_1 u_{t-1} + \varepsilon_t$

$$y_t = \delta_0 + \delta_1 t + \varphi_1 u_{t-1} + \varepsilon_t$$

But

$$y_{t-1} = \delta_0 + \delta_1 (t-1) + u_{t-1} \Rightarrow u_{t-1} = y_{t-1} - \delta_0 - \delta_1 (t-1)$$

$$y_t = \delta_0 + \delta_1 t + \varphi_1 [y_{t-1} - \delta_0 - \delta_1 (t-1)] + \varepsilon_t$$

$$y_t = \underbrace{[(1-\varphi_1)\delta_0 + \varphi_1 \delta_1]}_{\delta} + \underbrace{[(1-\varphi_1)\delta_1]}_{\gamma} t + \varphi_1 y_{t-1} + \varepsilon_t$$

$\rightarrow y_t = \varphi_1 y_{t-1} + \delta + \gamma t + \varepsilon_t$ H_1 in "constant and trend"-case

Verbeek (2000) "A Guide to Modern Econometrics"

Unit root and **AR lag order polynomial**

$$a(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0$$

$$a(1) = 1 - \varphi_1 1 - \varphi_2 1^2 - \ldots - \varphi_p 1^p = 0$$

$$\Rightarrow 1 = \varphi_1 + \varphi_2 + \ldots + \varphi_p$$

Example: Null hypotheses of **unit root tests**

$y_t = \varphi_1 y_{t-1} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$	$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t$
$H_0 : \varphi_1 = 1$	$H_0 : \varphi_1 + \varphi_2 = 1$	$H_0 : \varphi_1 + \varphi_2 + \ldots + \varphi_p = 1$

Testing for Unit Roots AR(2)

Augmented Dickey-Fuller test

with the same **asymptotic** critical values

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

Stationarity requires:

$$\varphi_2 + \varphi_1 < 1$$

$$\varphi_2 - \varphi_1 < 1$$

$$\varphi_2 > -1$$

Hypothesis:

$H_0 : \varphi_1 + \varphi_2 = 1$ given $\varphi_2 > -1$

Verbeek (2000) “A Guide to Modern Econometrics ”

Deriving the estimating equation for AR(2)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

$$= (\varphi_1 + \varphi_2) y_{t-1} - \varphi_2 (y_{t-1} - y_{t-2}) + \varepsilon_t$$

Hence, if we subtract y_{t-1} on both sides, we get

$$\Delta y_t = (\varphi_1 + \varphi_2 - 1) y_{t-1} - \varphi_2 \Delta y_{t-1} + \varepsilon_t$$

$$= \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \varepsilon_t$$

with $\pi_1 = \varphi_1 + \varphi_2 - 1$ and $\pi_2 = -\varphi_2$

$H_0: \varphi_1 + \varphi_2 = 1$ given $|\varphi_2| < 1$ is the same as

$H_0: \pi_1 = \varphi_1 + \varphi_2 - 1 = 0$

Hence, regress Δy_t on y_{t-1} and Δy_{t-1} and perform t-test based on $\hat{\pi}_1$ with DF-critical values given earlier.

Deriving the estimating equation for AR(3)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \varepsilon_t$$

$$= (\varphi_1 + \varphi_2 + \varphi_3) y_{t-1} - \varphi_2 (y_{t-1} - y_{t-2}) - \varphi_3 (y_{t-1} - y_{t-3}) + \varepsilon_t$$

Hence, if we subtract y_{t-1} on both sides, we get

$$\Delta y_t = (\varphi_1 + \varphi_2 + \varphi_3 - 1) y_{t-1} - \varphi_2 \Delta y_{t-1} - \varphi_3 (y_{t-1} - y_{t-3}) + \varepsilon_t$$

Hence, if we add and subtract $\varphi_3(y_{t-1}-y_{t-2})$, we get

$$\Delta y_t = (\varphi_1 + \varphi_2 + \varphi_3 - 1) y_{t-1} - \varphi_2 \Delta y_{t-1} - \varphi_3 \Delta y_{t-1}$$

$$- \varphi_3 (y_{t-1} - y_{t-3}) + \varphi_3 (y_{t-1} - y_{t-2}) + \varepsilon_t$$

$$= (\varphi_1 + \varphi_2 + \varphi_3 - 1) y_{t-1} - (\varphi_2 + \varphi_3) \Delta y_{t-1} - \varphi_3 \Delta y_{t-2}$$

$$= \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \pi_3 \Delta y_{t-2} + \varepsilon_t$$

with $\pi_1 = \varphi_1 + \varphi_2 + \varphi_3 - 1$; $\pi_2 = -(\varphi_2 + \varphi_3)$ and $\pi_3 = -\varphi_3$

Testing for (a single) Unit Root in an AR(*p*)

Rewriting works for any AR(p) process:

$$
\begin{aligned}
y_t &= \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t \\
&= (\varphi_1 + \varphi_2 + \ldots + \varphi_p) y_{t-1} - (\varphi_2 + \ldots + \varphi_p)(y_{t-1} - y_{t-2}) - \ldots \\
&\qquad - \varphi_p (y_{t-p+1} - y_{t-p}) + \varepsilon_t
\end{aligned}
$$

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \ldots + \pi_p \Delta y_{t-p+1} + \varepsilon_t$$

That is, $\pi_1 = \sum_{j=1}^{p} \varphi_j - 1$ and $\pi_i = -\sum_{j=i}^{p} \varphi_j$, for $i = 2,\ldots,p$

Verbeek (2000) “A Guide to Modern Econometrics ”

Testing for (a single) Unit Root in an AR(*p*)

Estimating equation:

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \ldots + \pi_p \Delta y_{t-p+1} + \varepsilon_t$$

Test of a single unit root is a test that $\pi_1 = 0$.

Regression of Δy_t on y_{t-1} and $\Delta y_{t-1},\ldots,\Delta_{t-p+1}$ and test the significance of y_{t-1}. Under the null of a single unit root, all variables are stationary, except y_{t-1}.

Testing for (a single) Unit Root in an ARMA process

Any ARMA model (with an invertible MA polynomial) can be written as an infinite autoregressive process.

That is, any unknown ARIMA(p, d, q) process can be well approximated by an ARIMA(n, d, 0) of order no more than $T^{1/3}$

(Said and Dickey (1984), Enders (1995), p.226)

So the above augmented regression can also be used to test for a unit root in an ARMA model.

Verbeek (2000) "A Guide to Modern Econometrics "

Which unit root test is adequate?

„Fit a specification that is a **plausible description of the data** under both the null and the alternative hypothesis.“

- for growing series: constant term plus linear trend

  exponentially growing series: take logs first $\log\left(e^{\delta t}\right) = \delta t$

- for non-growing series: constant, no trend regression

- not sure: use model with trend (first)

  If trend term is erroneously omitted, tests are biased toward unit roots; yet, including unnecessary trend also reduces power because coefficients eat up dof.

A word of caution

"Rejection will imply **no** unit root (with or without a deterministic trend). Because unit-root tests are notoriously lacking in power (i.e., they very often tell us there is a unit root when there is no unit root), this is usually taken as firm evidence against a unit root."

Elder and Kennedy (2001) Testing for Unit Roots

How do I choose the AR order of ADF test?

- Fit **several versions of Augmented DF-Tests** to allow for serial correlation of the errors

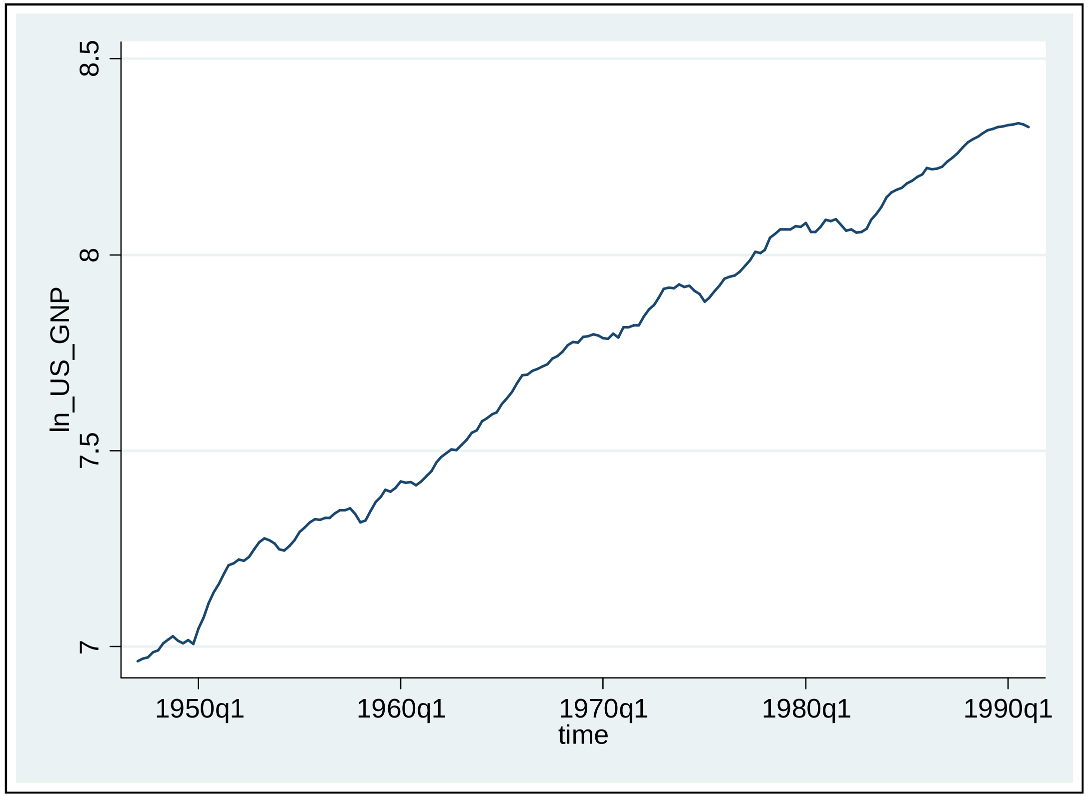
  or

- use a model selection criterion to determine the order of the regression, e.g. the **Hannan-Quinn criterion**

$$HQ(p) = \log \hat{\sigma}^2(p) + (1+p)\frac{2\ln(\ln(T))}{T}$$

Example: Quarterly US GNP (log)

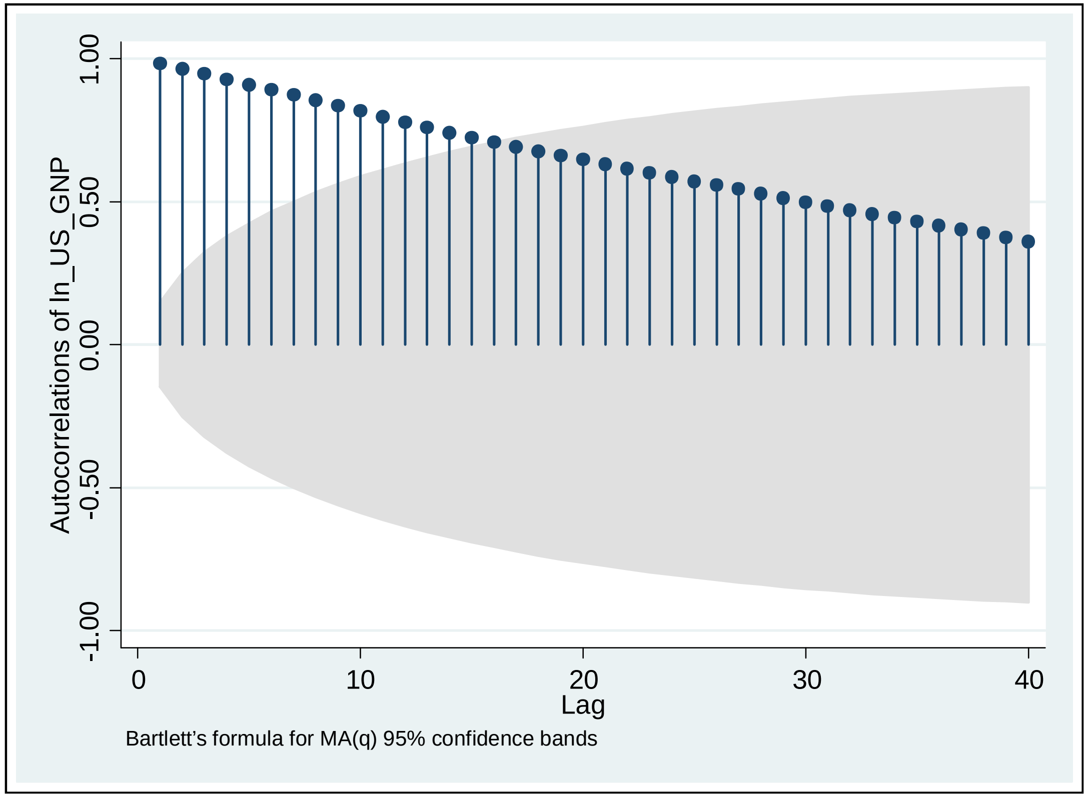
1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) "Time series Analysis and Its Applications"

Example: Quarterly US GNP (log)

Clear trend → **Case 3**: constant and trend

$$\boxed{\Delta y_t = \theta y_{t-1} + \delta + \gamma t + \varepsilon_t}$$

```
. regress D.ln_US_GNP L.ln_US_GNP time
      Source |       SS       df       MS              Number of obs =     176
-------------+------------------------------           F(  2,   173) =    2.76
       Model |  .000623477     2  .000311739           Prob > F      =  0.0659
    Residual |   .01951555   173  .000112807           R-squared     =  0.0310
-------------+------------------------------           Adj R-squared =  0.0198
       Total |  .020139027   175   .00011508           Root MSE      =  .01062

------------------------------------------------------------------------------
 D.ln_US_GNP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_US_GNP |
         L1. |  -.0302594   .0186748    -1.62   0.107    -.0671192    .0066004
        time |   .0002081   .0001459     1.43   0.155    -.0000798     .000496
       _cons |   .2334261   .1386819     1.68   0.094    -.0403004    .5071525
------------------------------------------------------------------------------
```

DF test statistic -1.62

Appropriate critical value at 5% is ~ -3.45 => no rejection of the null

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \delta + \gamma t + \varepsilon_t$$

Example: Quarterly US GNP (log)

Clear trend → **Case 3**: constant and trend

```
. regress D.ln_US_GNP L.ln_US_GNP D.L.ln_US_GNP time
      Source |       SS       df       MS              Number of obs =     175
-------------+------------------------------           F(  3,   171) =   12.47
       Model |  .003615645     3  .001205215           Prob > F      =  0.0000
    Residual |   .01652135   171  .000096616           R-squared     =  0.1796
-------------+------------------------------           Adj R-squared =  0.1652
       Total |  .020136996   174   .00011573           Root MSE      =  .00983

------------------------------------------------------------------------------
 D.ln_US_GNP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_US_GNP |
         L1. |  -.0427629   .0175261    -2.44   0.016    -.0773582   -.0081676
         LD. |   .3905358   .0705662     5.53   0.000     .2512428    .5298288
        time |   .0003134   .0001369     2.29   0.023     .0000433    .0005836
       _cons |   .3229541   .1300965     2.48   0.014     .0661522    .5797561
------------------------------------------------------------------------------
```

DF test statistic -2.44 > -3.45 => no rejection of the null

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. dfuller ln_US_GNP, trend
Dickey-Fuller test for unit root                   Number of obs   =       176

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.620            -4.015            -3.440            -3.140
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7844

. dfuller ln_US_GNP, trend lags(1)
Augmented Dickey-Fuller test for unit root         Number of obs   =       175

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.440            -4.015            -3.440            -3.140
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3586
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. dfuller ln_US_GNP, trend lags(2)
Augmented Dickey-Fuller test for unit root         Number of obs   =       174

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.892            -4.015            -3.440            -3.140
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1649

. dfuller ln_US_GNP, trend lags(3)
Augmented Dickey-Fuller test for unit root         Number of obs   =       173

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.503            -4.016            -3.440            -3.140
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3262
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. dfuller ln_US_GNP, trend lags(4)
Augmented Dickey-Fuller test for unit root         Number of obs   =       172

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.366            -4.016            -3.440            -3.140
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3979

. dfuller ln_US_GNP, trend lags(5)
Augmented Dickey-Fuller test for unit root         Number of obs   =       171

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.246            -4.016            -3.441            -3.141
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.4638
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. dfuller ln_US_GNP, trend lags(6)
Augmented Dickey-Fuller test for unit root         Number of obs   =       170

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.384            -4.017            -3.441            -3.141
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3882
```

Summary:

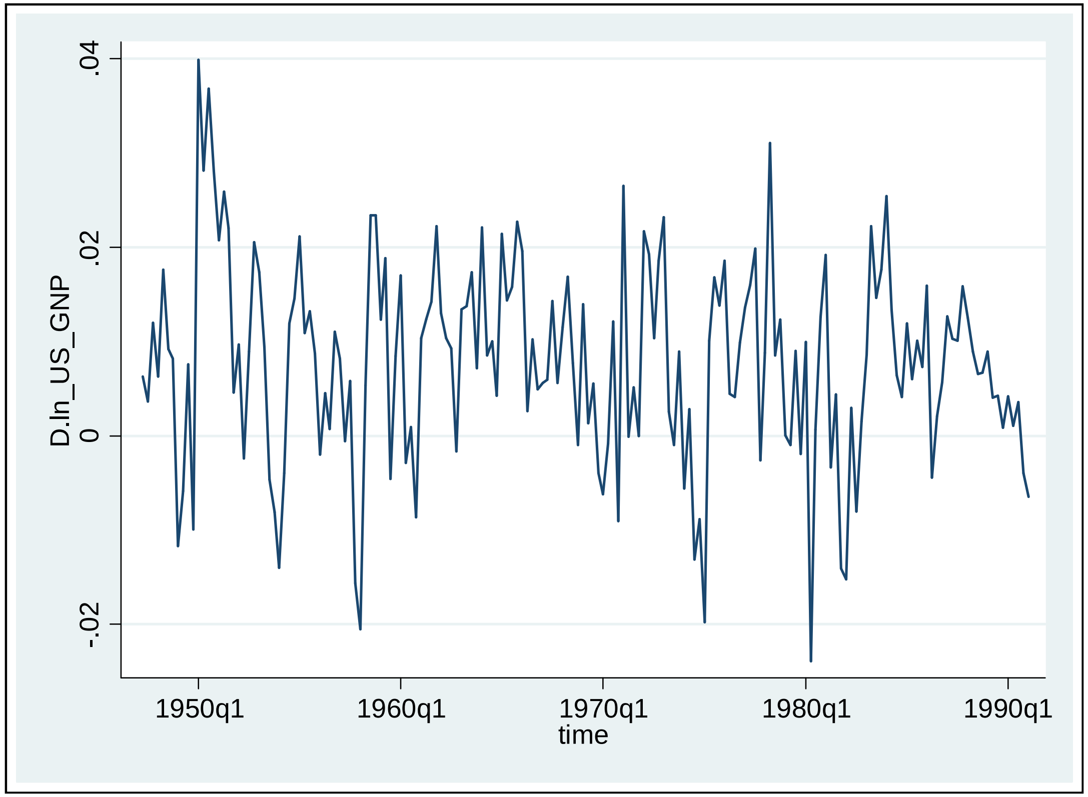
DF	ADF(1)	ADF(2)	ADF(3)	ADF(4)	ADF(5)	ADF(6)
-1.62	-2.44	-2.89	-2.5	-2.37	-2.54	-2.38

→ the conclusion does not change, and **we cannot reject the presence of a first unit root**

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log) – **First Difference**

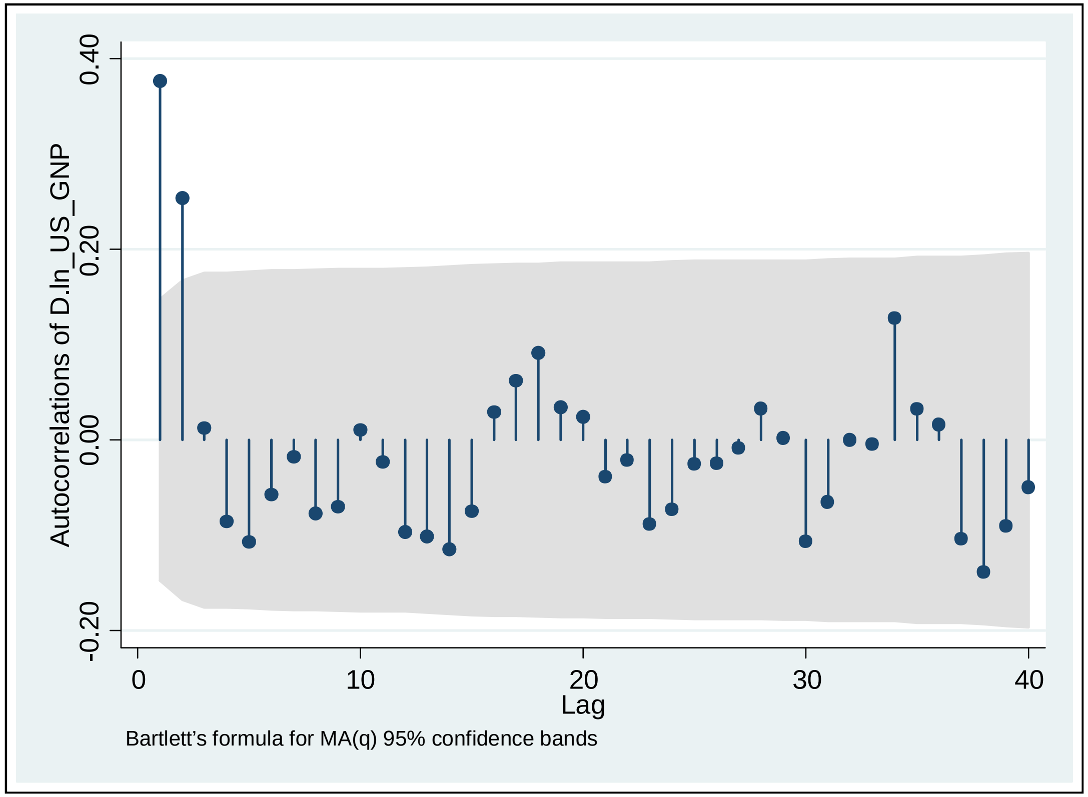
1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) "Time series Analysis and Its Applications"

Example: Quarterly US GNP (log) – **First Difference**

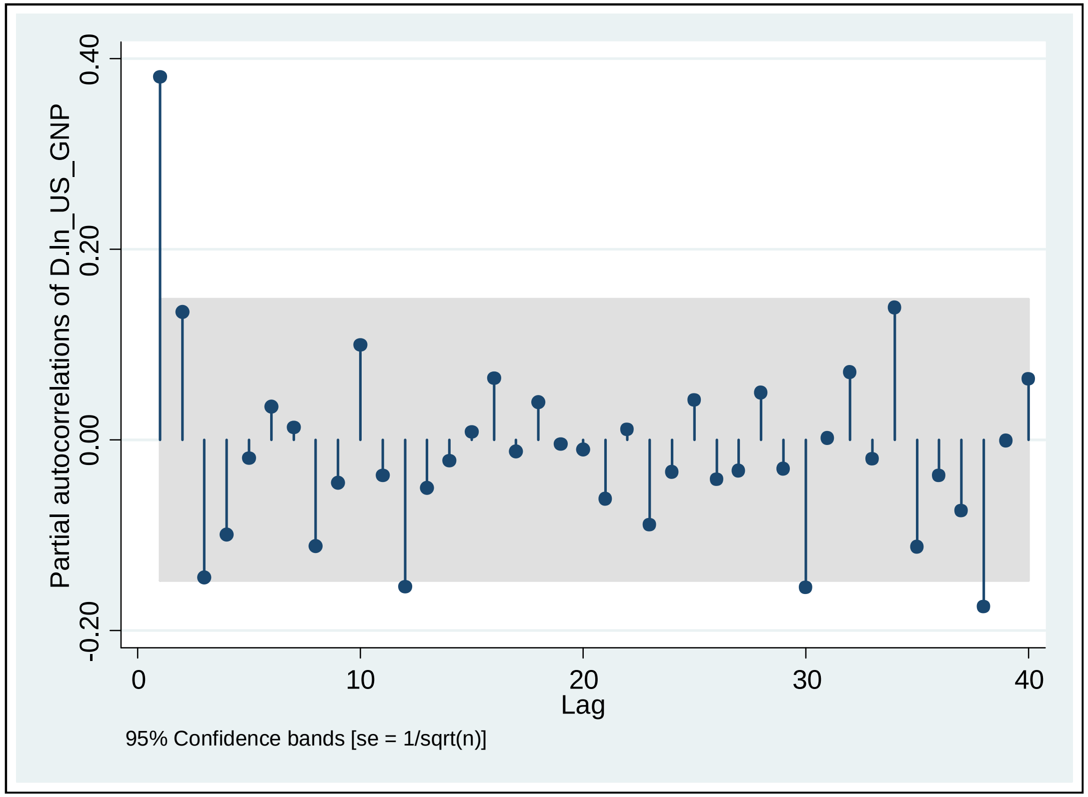
1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) "Time series Analysis and Its Applications"

Example: Quarterly US GNP (log) – **First Difference**

1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. arima D.ln_US_GNP, ma(1/2)
[...]
ARIMA regression

Sample:  1947q2 to 1991q1                       Number of obs      =       176
                                                Wald chi2(2)       =     26.48
Log likelihood =  565.1442                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
 D.ln_US_GNP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_US_GNP    |
       _cons |   .0076825   .0011899     6.46   0.000     .0053503    .0100147
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3122641   .0673827     4.63   0.000     .1801964    .4443318
         L2. |   .2713829   .0701125     3.87   0.000      .133965    .4088008
-------------+----------------------------------------------------------------
      /sigma |   .0097483   .0004685    20.81   0.000       .00883    .0106667
------------------------------------------------------------------------------
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. arima ln_US_GNP, arima(0,1,2)
[...]
ARIMA regression

Sample:  1947q2 to 1991q1                        Number of obs      =       176
                                                 Wald chi2(2)       =     26.48
Log likelihood =  565.1442                       Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
 D.ln_US_GNP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_US_GNP    |
       _cons |   .0076825   .0011899     6.46   0.000     .0053503    .0100147
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3122641   .0673827     4.63   0.000     .1801964    .4443318
         L2. |   .2713829   .0701125     3.87   0.000      .133965    .4088008
-------------+----------------------------------------------------------------
      /sigma |   .0097483   .0004685    20.81   0.000       .00883    .0106667
------------------------------------------------------------------------------
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
------------------------------------------------------------------------------
             |                 OPG
 D.ln_US_GNP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_US_GNP    |
       _cons |   .0076825   .0011899     6.46   0.000     .0053503    .0100147
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3122641   .0673827     4.63   0.000     .1801964    .4443318
         L2. |   .2713829   .0701125     3.87   0.000      .133965    .4088008
-------------+----------------------------------------------------------------
      /sigma |   .0097483   .0004685    20.81   0.000       .00883    .0106667
------------------------------------------------------------------------------
```

$$x_t = 0.0077 + \varepsilon_t + 0.3122\varepsilon_{t-1} + 0.2714\varepsilon_{t-2}$$

$x_t \sim$ **ARMA(0,2)**

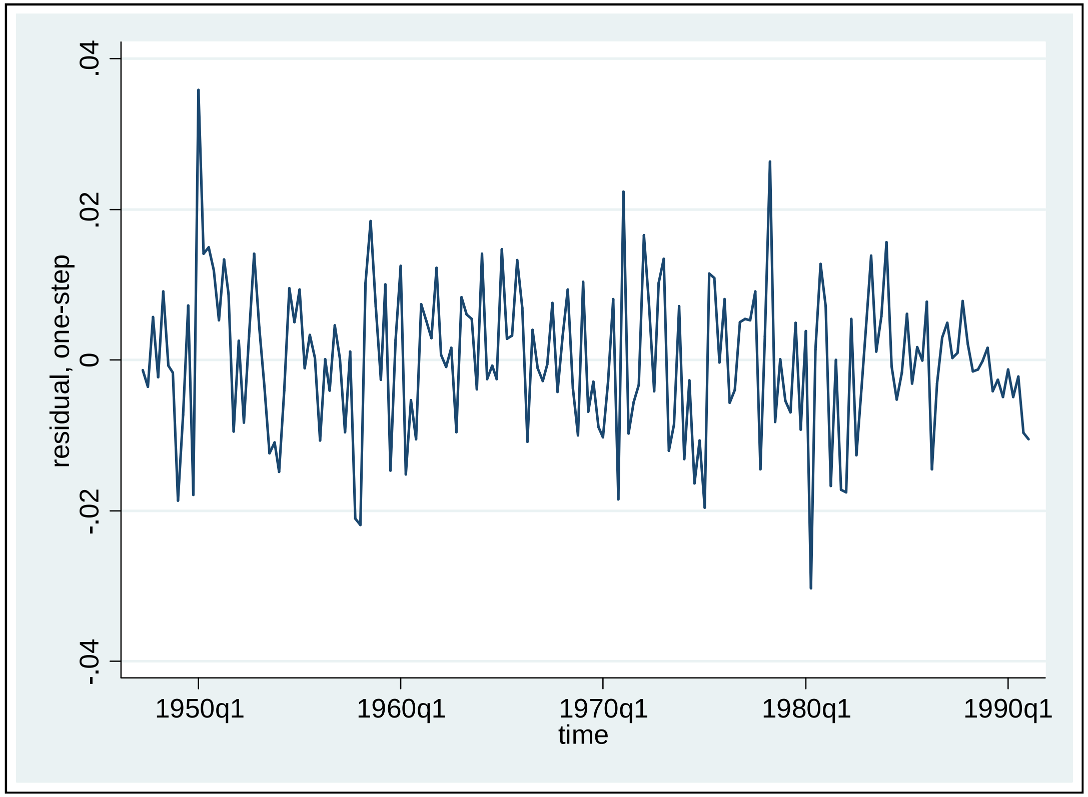
$$(1-L)y_t = 0.0077 + \varepsilon_t + 0.3122\varepsilon_{t-1} + 0.2714\varepsilon_{t-2}$$

$y_t \sim$ **ARIMA(0,1,2)**

Shumway, Stoffer (2000) "Time series Analysis and Its Applications"

Example: Quarterly US GNP (log) – **First Difference**

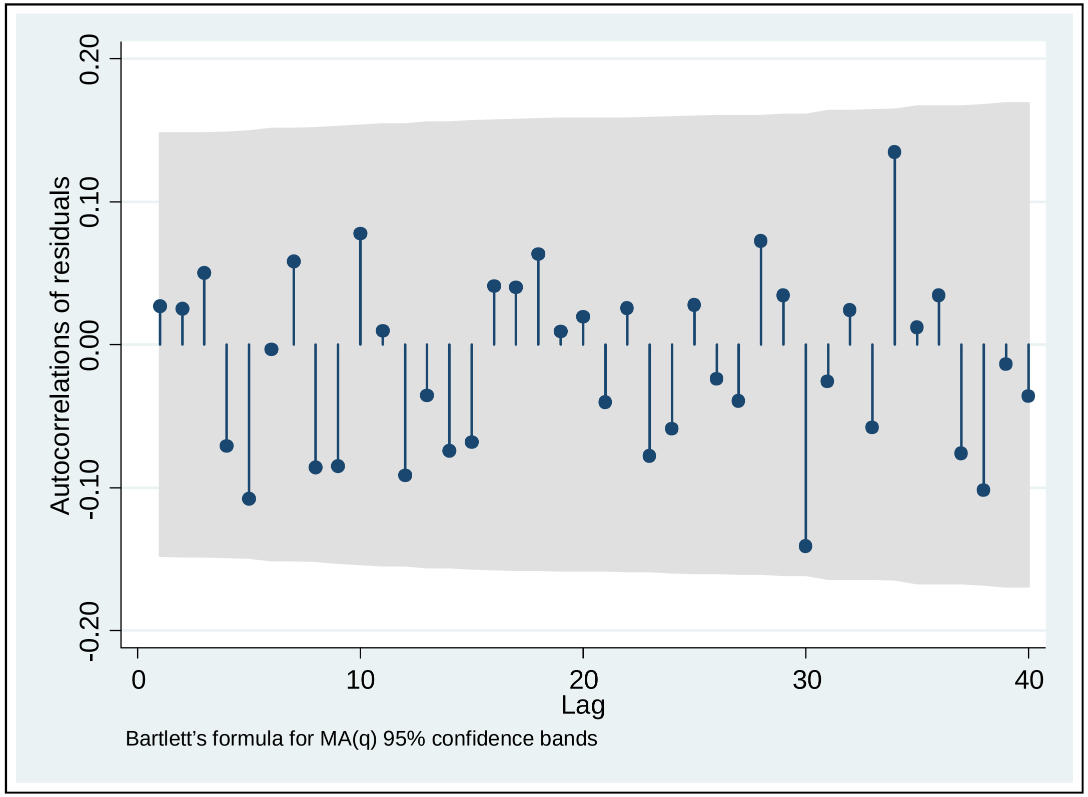
1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log) – **First Difference**

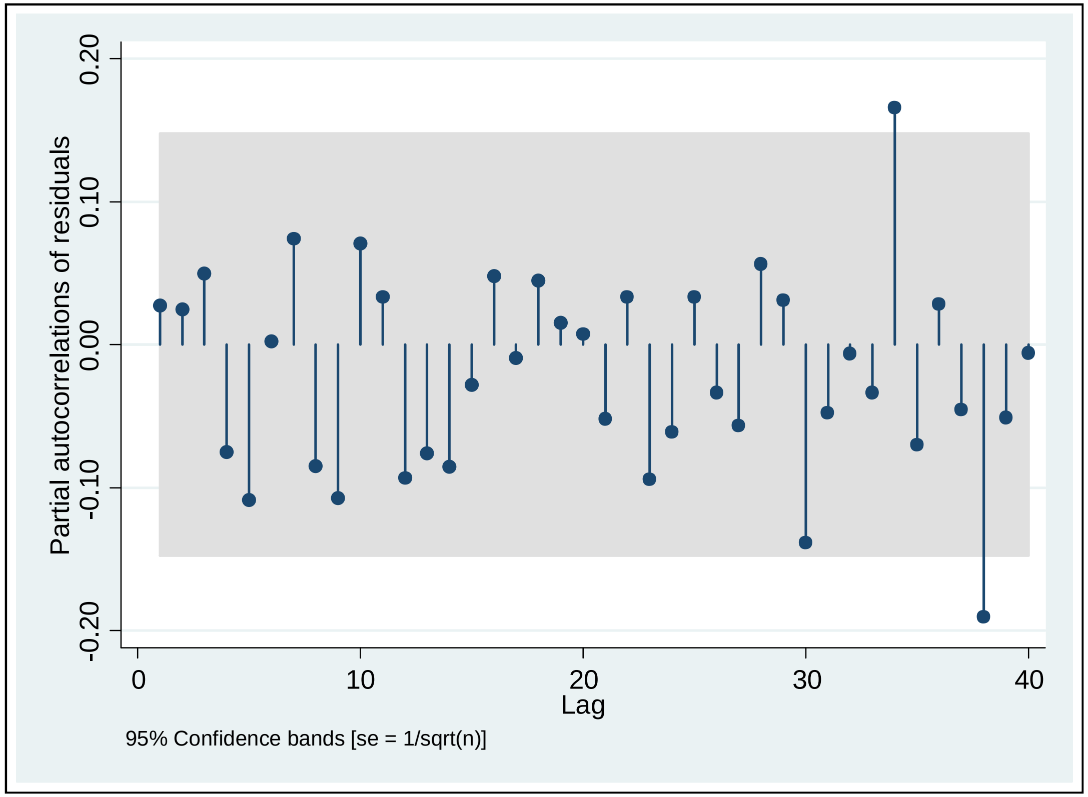
1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) "Time series Analysis and Its Applications"

Example: Quarterly US GNP (log) – First Difference

1947q1-1991q1 seasonally adjusted



Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

Example: Quarterly US GNP (log)

```
. corrgram res
                                               -1       0       1 -1       0       1
 LAG       AC       PAC       Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.0272   0.0274   .13218  0.7162           |                  |
2        0.0252   0.0249   .24678  0.8839           |                  |
3        0.0501   0.0497   .70185  0.8728           |                  |
4       -0.0712  -0.0752   1.6246  0.8044           |                  |
5       -0.1076  -0.1085   3.7476  0.5863           |                  |
6       -0.0033   0.0025   3.7496  0.7105           |                  |
[...]
30      -0.1408  -0.1387   22.228  0.8456          -|                 -|
31      -0.0257  -0.0477   22.371  0.8710           |                  |
32       0.0244  -0.0061   22.501  0.8934           |                  |
33      -0.0580  -0.0332   23.237  0.8965           |                  |
34       0.1351   0.1659   27.262  0.7870           |-                 |-
35       0.0121  -0.0701   27.294  0.8206           |                  |
36       0.0344   0.0284   27.559  0.8425           |                  |
37      -0.0761  -0.0453   28.866  0.8280           |                  |
38      -0.1017  -0.1907   31.214  0.7741           |                 -|
39      -0.0138  -0.0509   31.258  0.8066           |                  |
40      -0.0361  -0.0058   31.559  0.8273           |                  |
```

Shumway, Stoffer (2000) “Time series Analysis and Its Applications”

DS-Model

$$\Delta y_t = 0.003 + 0.369\Delta y_{t-1} + \varepsilon_t$$

$$(1 - 0.369L)\Delta y_t = 0.003 + \varepsilon_t$$

$$y_t = 1.369 y_{t-1} - 0.369 y_{t-2} + 0.003 + \varepsilon_t$$

TS-Model

$$y_t = -0.321 + 0.0003t + 1.335 y_{t-1} - 0.401 y_{t-2} + \varepsilon_t$$

Unit root test for AR(2) (with trend)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \delta + \gamma t + \varepsilon_t$$

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2 \Delta y_{t-1} + \delta + \gamma t + \varepsilon_t$$

$\pi_1 = \varphi_1 + \varphi_2 - 1$ and $\pi_2 = -\varphi_2$

$H_0: \varphi_1 + \varphi_2 = 1$ or

$H_0: \pi_1 = \varphi_1 + \varphi_2 - 1 = 0$

$$Y_t = \hat{\mu} + \hat{\gamma} t + \hat{\delta} Y_{t-1} + \hat{\phi}_1 \Delta Y_{t-1} + \hat{\varepsilon}_t .$$

The augmented Dickey-Fuller unit-root test (David A. Dickey and Wayne F. Fuller, 1981) is often used to try to distinguish a TS model from a DS model.[7] For the second-order models under consideration, the augmented Dickey-Fuller regression takes the following form:

$$(4) \quad Y_t = \hat{\mu} + \hat{\gamma} t + \hat{\delta} Y_{t-1} + \hat{\phi}_1 \Delta Y_{t-1} + \hat{\varepsilon}_t .$$

Under the unit-root (or DS model) null hypothesis, $\delta = 1$; thus, the Dickey-Fuller test statistic is simply the t test, $\hat{\tau} \equiv (\hat{\delta} - 1)/\mathrm{SE}(\hat{\delta})$, where $\mathrm{SE}(\hat{\delta})$ is the standard error of the estimated coefficient.

For the postwar real GNP data under consideration, the sample value of the Dickey-Fuller test, which is denoted as $\hat{\tau}_{samp}$, is equal to -2.98. However, this statistic does not have the usual Student-*t* distribution, but is skewed toward negative values. At the 10-percent significance level, Dickey and Fuller (1981) calculate the appropriate asymptotic critical value to be -3.12. Thus, the evidence from this sample, in accordance with the findings of previous researchers, suggests that the DS model for real GNP cannot be rejected at even the 10-percent level.

However, the critical values provided by Dickey and Fuller (1981) for their augmented test are only valid asymptotically. In finite samples, the distribution of $\hat{\tau}$ will usually depend on the sample size and nuisance-parameter values (see e.g., Gene Evans and Savin, 1984). These factors can be taken into account by examining simulated data from the DS_{OLS} model and calculating the exact probability of obtaining the sample value of the test statistic from this particular null model. This ensures correct size for the test. More importantly, however, by simulating the TS_{OLS} model, the exact probability of obtaining $\hat{\tau}_{samp}$ from this particular alternative model can also be obtained. This allows correct assessment of test power against what is arguably one of the most interesting alternatives.

The DS_{OLS} model

$$y_t = 1369 y_{t-1} - 0.369 y_{t-2} + 0.003 + \varepsilon_t$$

10000 Time series are generated according to this model by simulating independent normal errors.
Each time parameters of $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-1} + \delta + \varepsilon_t$
are estimated by OLS and DF statistic is calculated.
Density of DF values is on next slide

Similarly for the TS_{OLS} model

$$y_t = -0.321 + 0.0003t + 1.335 y_{t-1} - 0.401 y_{t-2} + \varepsilon_t$$

The Uncertain Unit Root in Real GNP

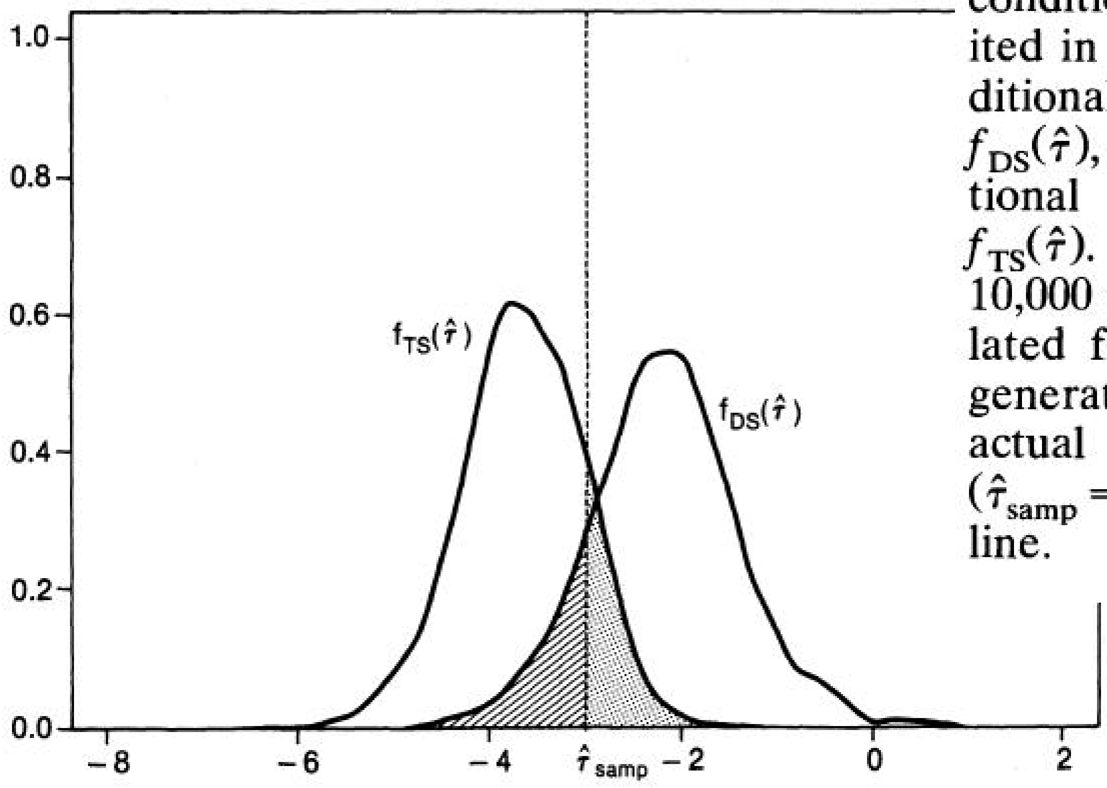


FIGURE 1. EMPIRICAL DENSITIES OF $\hat{\tau}$ FROM DS_{OLS} AND TS_{OLS} MODELS

The test-statistic probability distributions conditional on the OLS models are exhibited in Figure 1. The distribution of $\hat{\tau}$ conditional on the DS_{OLS} model is denoted $f_{DS}(\hat{\tau})$, while the distribution of $\hat{\tau}$ conditional on the TS_{OLS} model is denoted $f_{TS}(\hat{\tau})$. Each distribution is formed from 10,000 realizations of the test statistic calculated from 10,000 simulated data samples generated from the particular model.[8] The actual sample value of the test statistic ($\hat{\tau}_{samp} = -2.98$) is shown as a vertical dotted line.

[8]The samples are generated with normal independently and identically distributed errors with sample size and initial conditions that matched those in equations (1) and (2).

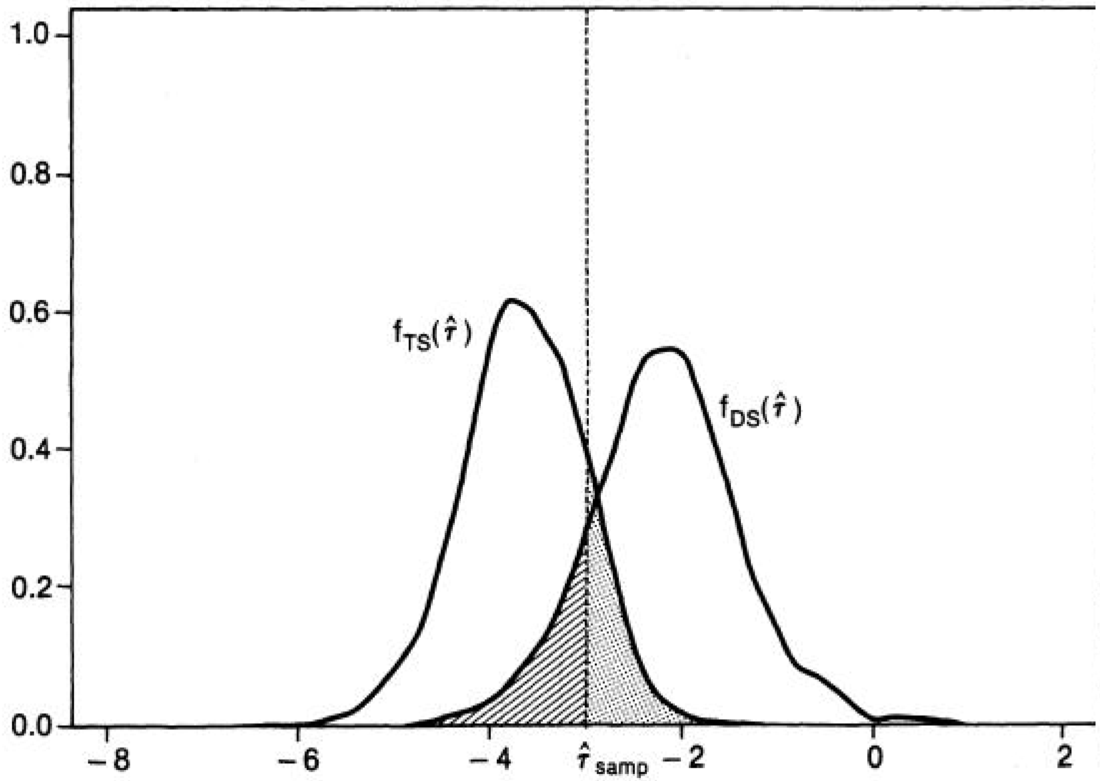


FIGURE 1. EMPIRICAL DENSITIES OF $\hat{\tau}$ FROM DS_{OLS} AND TS_{OLS} MODELS

There are two areas in Figure 1 of special interest. The hatched area under $f_{DS}(\hat{\tau})$ and to the left of $\hat{\tau}_{samp}$ represents the probability of obtaining a value of the t test equal to or smaller than −2.98, conditional on the DS model of equation (2). This p value, is denoted as

$$DS_{OLS}\ p \text{ value}$$

$$\equiv \operatorname{prob}\left(\hat{\tau} \le \hat{\tau}_{samp} \mid DS_{OLS} \text{ model}\right)$$

and represents the marginal significance level for rejection of the null hypothesis for the DS_{OLS} model. This probability equals 0.15; that is, given the sample test statistic, one could not reject the DS model at anything less than the 15-percent level in a classical hypothesis test. This is consistent with the usual inability to reject the DS model for real GNP at conventional significance levels.

Application: The Uncertain Unit Root in Real GNP

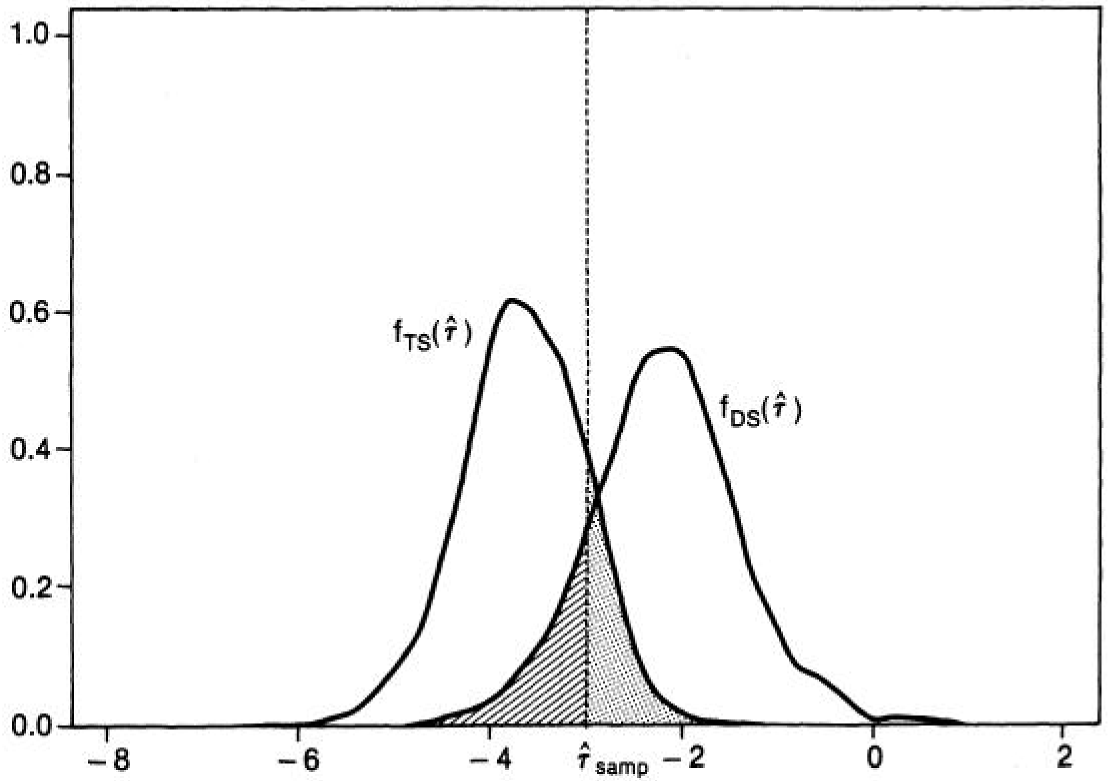


FIGURE 1. EMPIRICAL DENSITIES OF $\hat{\tau}$ FROM DS_{OLS} AND TS_{OLS} MODELS

The other area of interest is the shaded region under $f_{TS}(\hat{\tau})$ and to the right of $\hat{\tau}_{samp}$. This area represents the probability of obtaining a value of the t test equal to or greater than −2.98, conditional on the TS model of equation (1). This probability is denoted as

$$TS_{OLS}\ p\ \text{value}$$

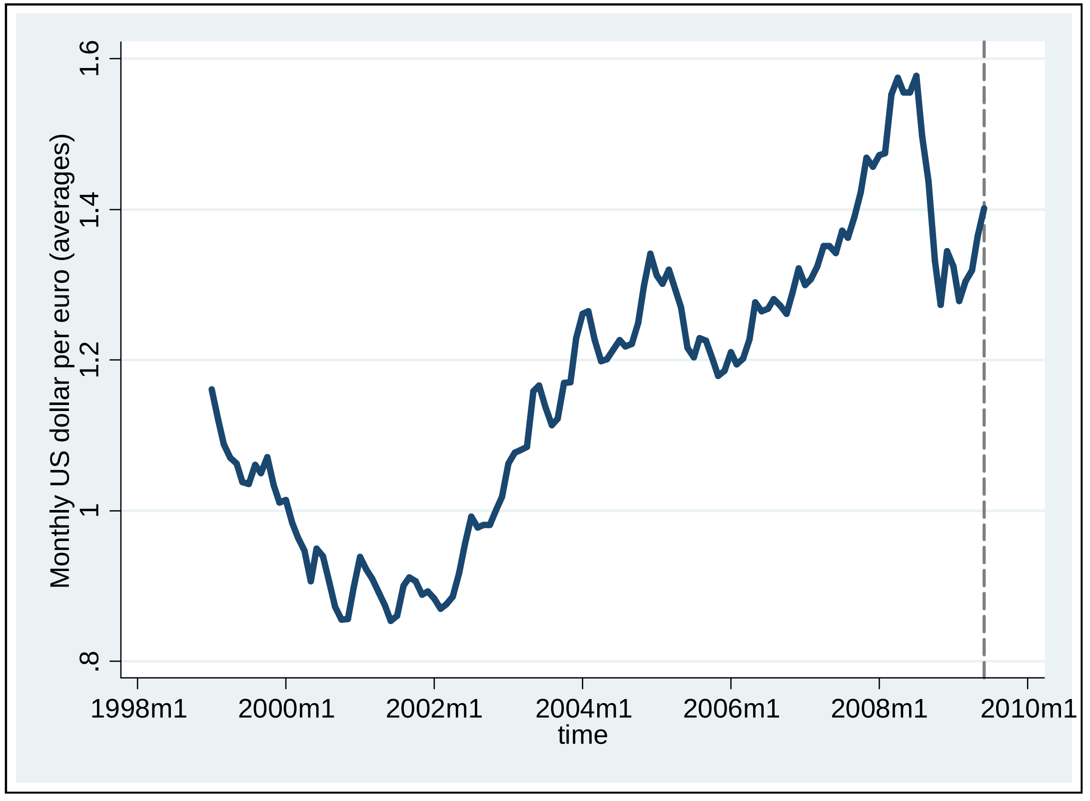
$$\equiv \text{prob}\left(\hat{\tau} \geq \hat{\tau}_{samp} \mid TS_{OLS}\ \text{model}\right).$$

For real GNP, the TS_{OLS} p value is 0.22, so one would not be able to reject the estimated TS_{OLS} model at even the 20-percent significance level.[9]

In short, the sample statistic for the augmented Dickey-Fuller test does not provide strong evidence against either the estimated DS_{OLS} model or the TS_{OLS} model for real GNP. Earlier papers that are unable to

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Example: US dollar per euro (exchange rate)

No clear trend → **Case 2**: constant, no trend

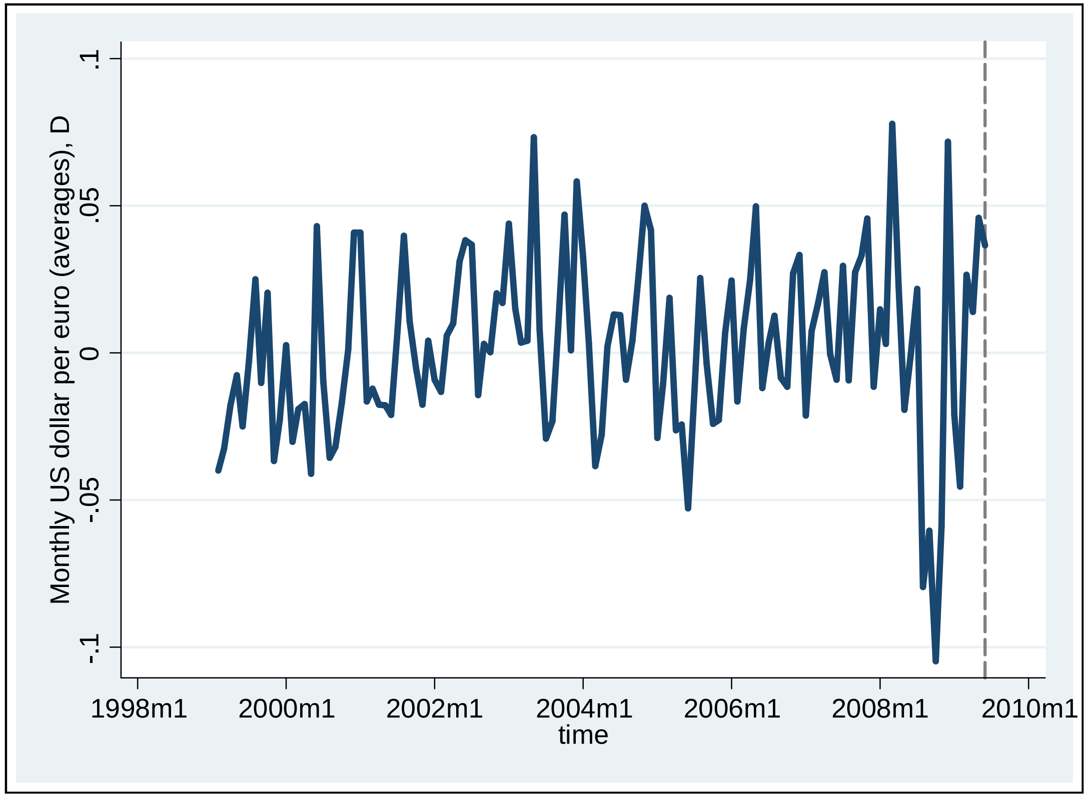
```
. dfuller exchange_rate if time <= 593,
Dickey-Fuller test for unit root                   Number of obs   =       125

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.435            -3.502            -2.888            -2.578
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9041
```

DF	ADF(1)	ADF(2)	ADF(3)	ADF(4)	ADF(5)	ADF(6)
-0.435	-0.842	-0.712	-0.774	-0.810	-0.899	-0.757

→ the conclusion does not change, and **we cannot reject the presence of a first unit root**

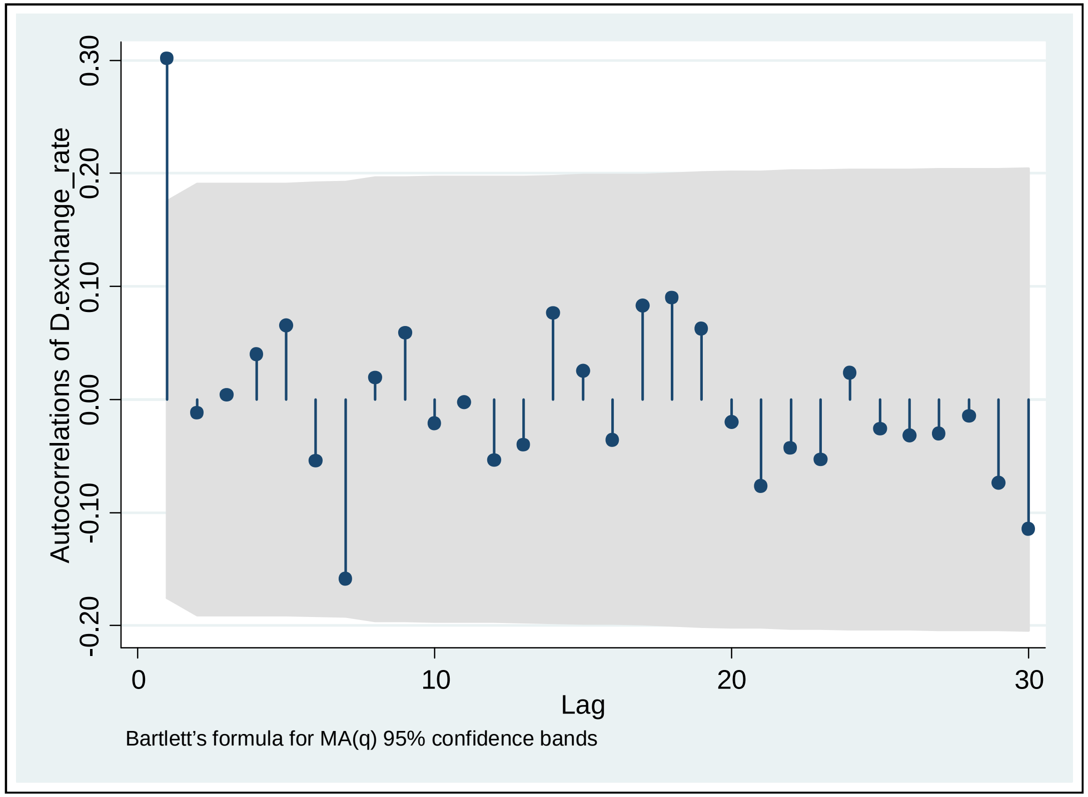
Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)



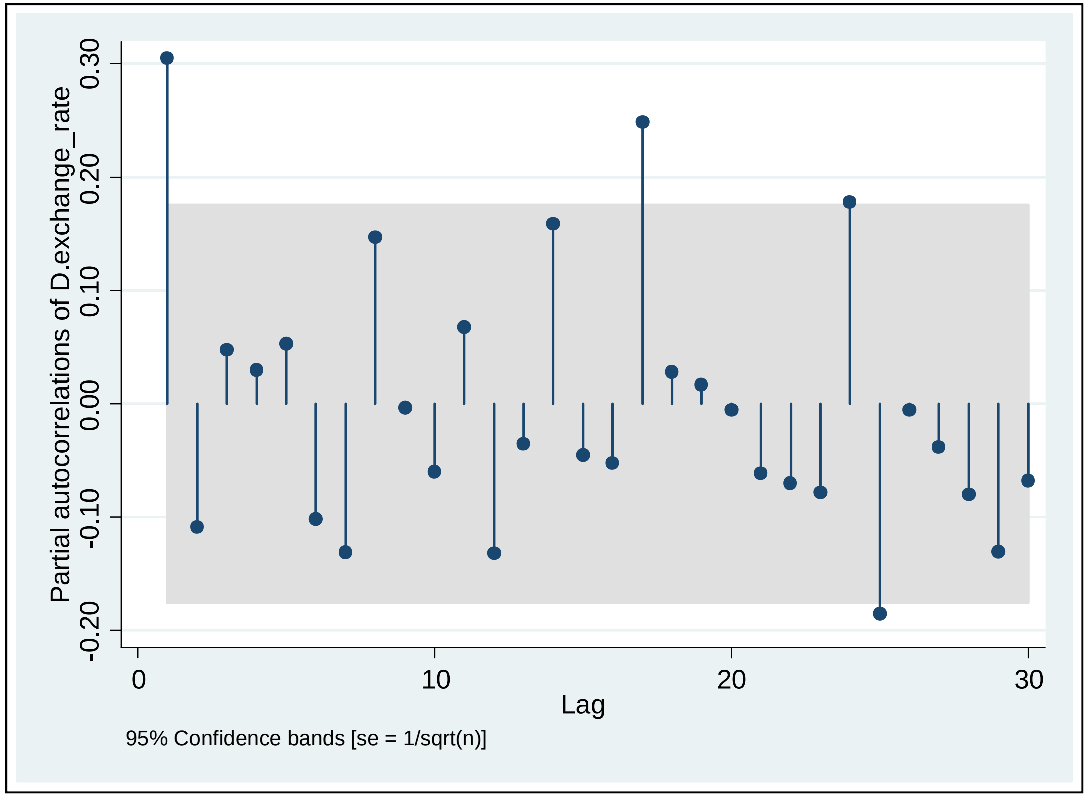
Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)



Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)



Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(0,1,1)
[...]
ARIMA regression

Sample:  1999m2 to 2009m6                       Number of obs      =        125
                                                Wald chi2(1)       =      12.87
Log likelihood =  266.9772                      Prob > chi2        =     0.0003

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018907   .0034256     0.55   0.581    -.0048232    .0086047
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3452627    .096229     3.59   0.000     .1566573    .5338682
-------------+----------------------------------------------------------------
      /sigma |    .028576   .0015061    18.97   0.000     .0256241     .031528
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(1,1,0)
[...]
ARIMA regression

Sample:  1999m2 to 2009m6                          Number of obs      =       125
                                                   Wald chi2(1)       =     17.43
Log likelihood =  266.1319                         Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018981    .003743     0.51   0.612     -.005438    .0092343
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3067845    .073477     4.18   0.000     .1627723    .4507968
-------------+----------------------------------------------------------------
      /sigma |   .0287709   .0015431    18.65   0.000     .0257465    .0317953
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(1,1,1)
[...]
ARIMA regression
Sample:  1999m2 to 2009m6                        Number of obs      =       125
                                                 Wald chi2(2)       =     12.87
Log likelihood =  266.9795                       Prob > chi2        =    0.0016
------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018907   .0034643     0.55   0.585    -.0048992    .0086806
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.0180361   .1774111    -0.10   0.919    -.3657554    .3296832
          ma |
         L1. |   .3611031   .1877277     1.92   0.054    -.0068364    .7290426
-------------+----------------------------------------------------------------
      /sigma |   .0285736   .0015466    18.48   0.000     .0255423    .0316049
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

Model	ARIMA(0,1,1)	ARIMA(1,1,0)	ARIMA(1,1,1)
AIC	-7.094	-7.081	-7.079
BIC	-7.033	-7.020	-6.956

$x_t \sim$ **ARMA(0,1)**

$$x_t = \varepsilon_t + 0.3471\varepsilon_{t-1}$$

$y_t \sim$ **ARIMA(0,1,1)**

$$(1-L)y_t = \varepsilon_t + 0.3471\varepsilon_{t-1}$$

Recall:

$$AIC = \log(\hat{\sigma}^2) + 2\frac{p+q}{T}$$

$$BIC = \log(\hat{\sigma}^2) + 2\frac{p+q}{T}\log T$$

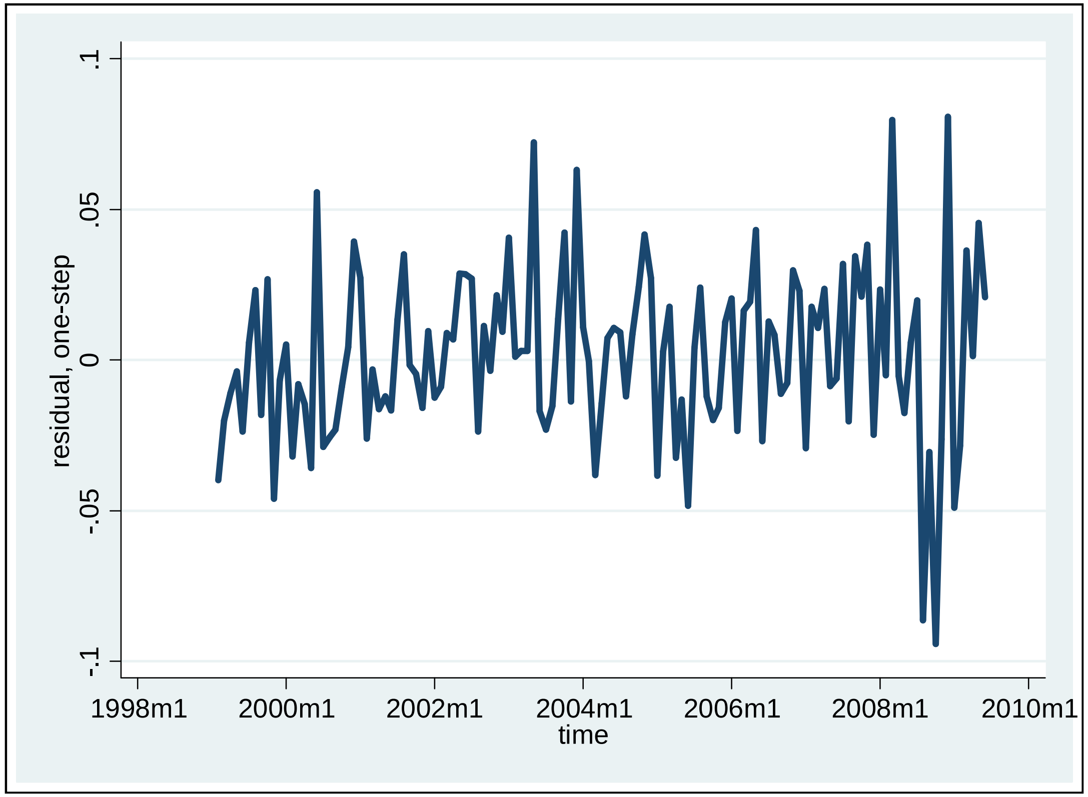
For both criteria:
the smaller the better!

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(0,1,1) noconst

[...]

ARIMA regression

Sample:  1999m2 to 2009m6                       Number of obs      =       125
                                                Wald chi2(1)       =     12.48
Log likelihood =  266.8257                      Prob > chi2        =    0.0004

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3470594   .0982231     3.53   0.000     .1545457    .5395731
-------------+----------------------------------------------------------------
      /sigma |   .0286087   .0015099    18.95   0.000     .0256493     .031568
------------------------------------------------------------------------------
```

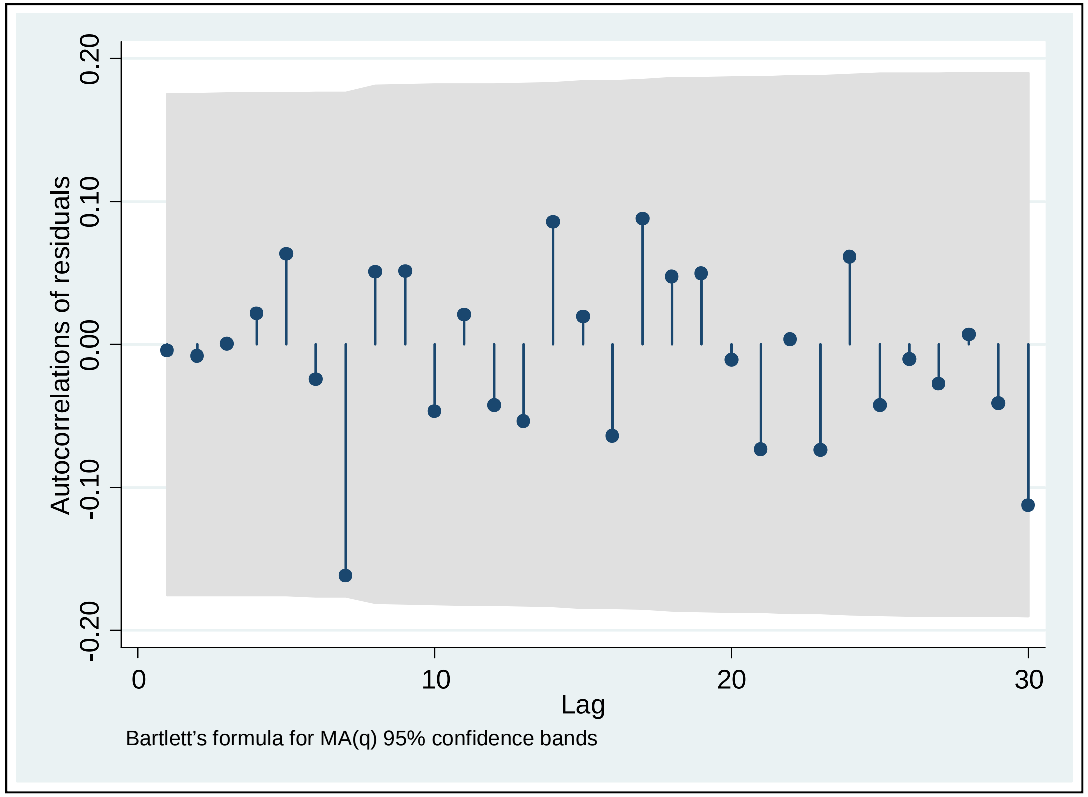
Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)



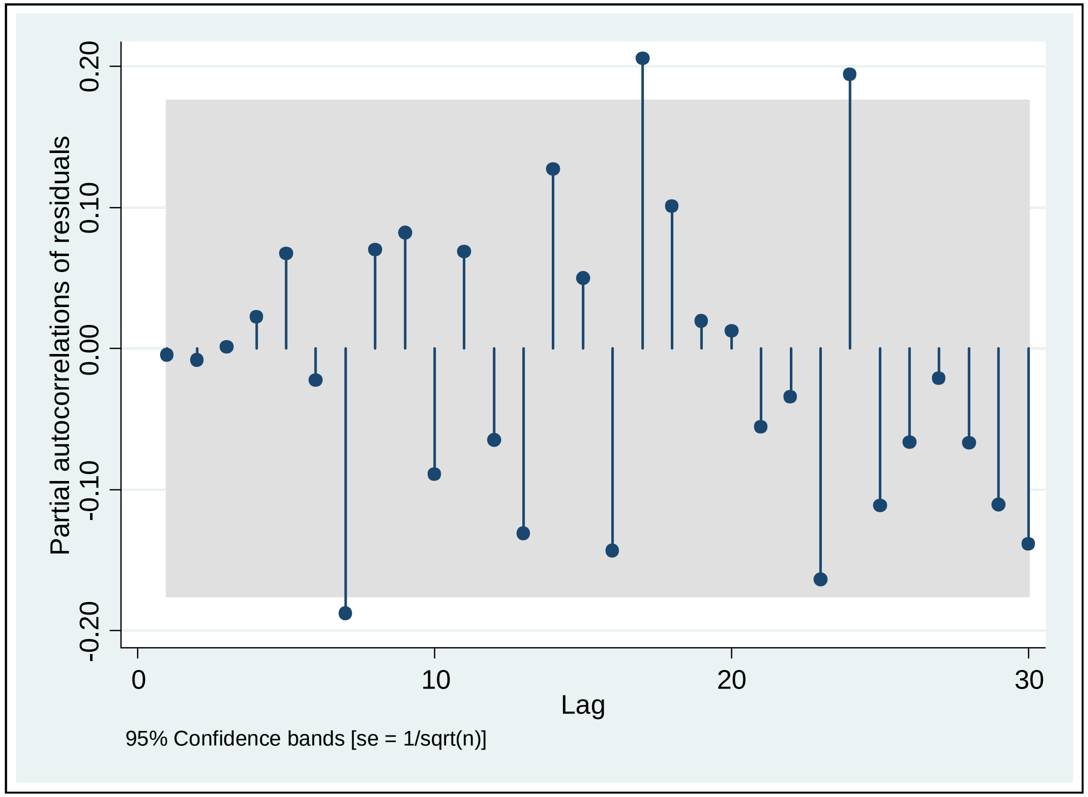
Example: US dollar per euro – First Difference

ACF of residuals



Example: US dollar per euro – First Difference

PACF of residuals



Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)

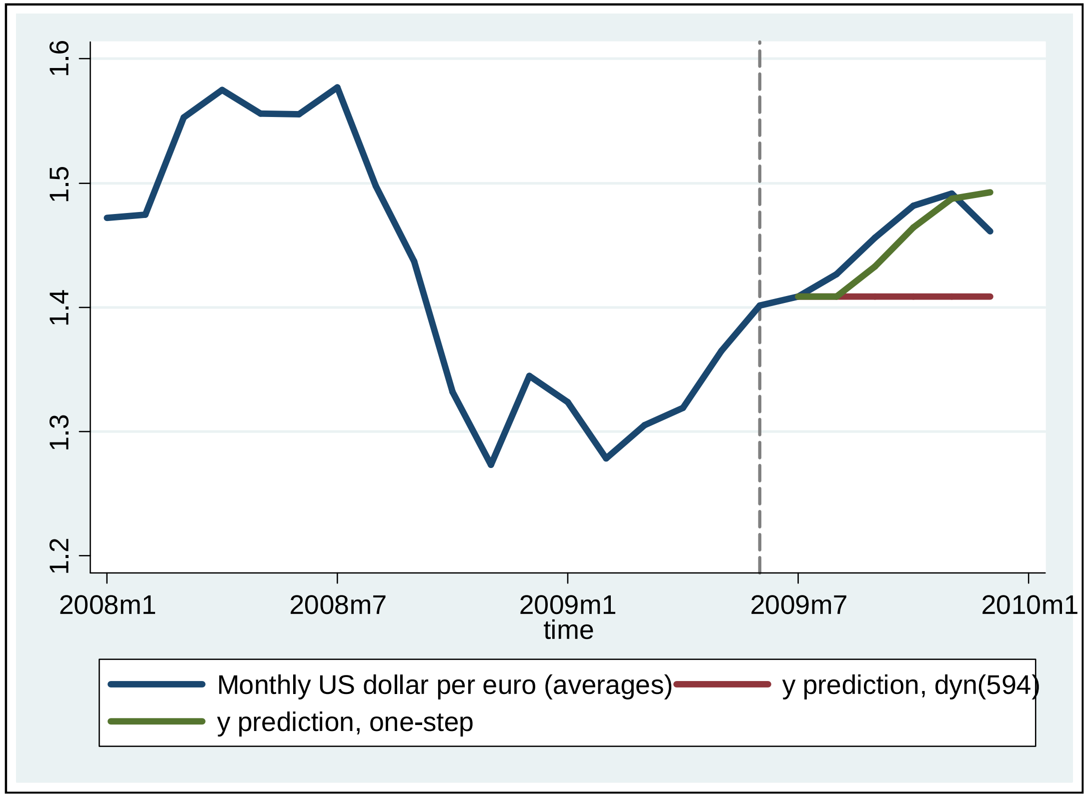
Recall:

$$Q = T(T+2)\sum_{k=1}^{K}\frac{1}{T-K}\hat{\rho}_k^2$$

```
. corrgram residuals, lags(30)
                                                 -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        -0.0044  -0.0044   .0025  0.9601          |                   |
2        -0.0080  -0.0081  .01085  0.9946          |                   |
3         0.0008   0.0009  .01093  0.9997          |                   |
4         0.0214   0.0224  .07111  0.9994          |                   |

[...]

19        0.0500   0.0196  9.4702  0.9648          |                   |
20       -0.0105   0.0125  9.4867  0.9766          |                   |
21       -0.0731  -0.0556  10.302  0.9747          |                   |
22        0.0038  -0.0346  10.304  0.9833          |                   |
23       -0.0738  -0.1635  11.152  0.9816          |                  -|
24        0.0612   0.1939  11.741  0.9827          |                   |-
25       -0.0422  -0.1114  12.023  0.9864          |                   |
26       -0.0102  -0.0666   12.04  0.9909          |                   |
27       -0.0275  -0.0214  12.162  0.9936          |                   |
28        0.0068  -0.0671  12.169  0.9959          |                   |
29       -0.0408  -0.1104  12.445  0.9969          |                   |
30       -0.1128  -0.1389  14.572  0.9919          |                  -|
```

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Recall

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T$$

$$\tilde{y}_{T+l|T} = \tilde{y}_{T+1|T}, \quad l = 2,3,\ldots$$

$$\tilde{y}_{T+1|T} = \lambda y_T + (1-\lambda)\tilde{y}_{T|T-1}$$

with $\lambda = (1+\theta_1)$ See below

Here

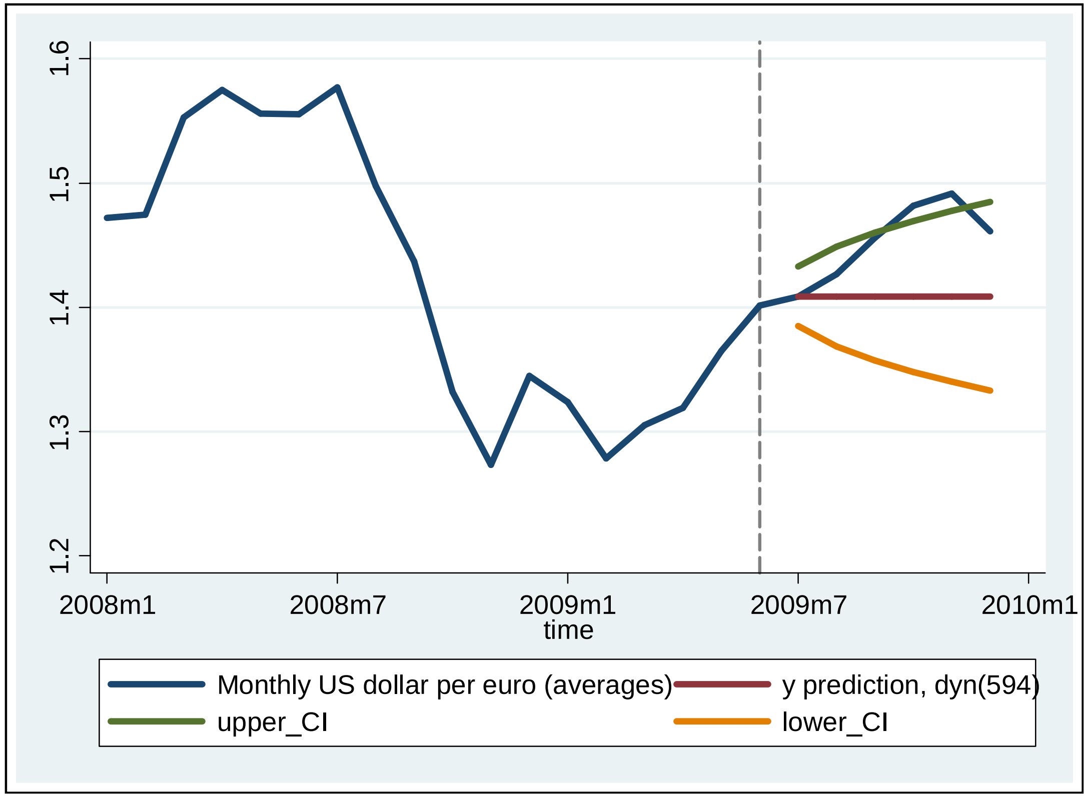
$$y_t = y_{t-1} + \varepsilon_t + 0.3471\varepsilon_{t-1}$$

and $\hat{\theta} =$ -0.3471

$$\hat{\lambda} = 0.6529$$

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



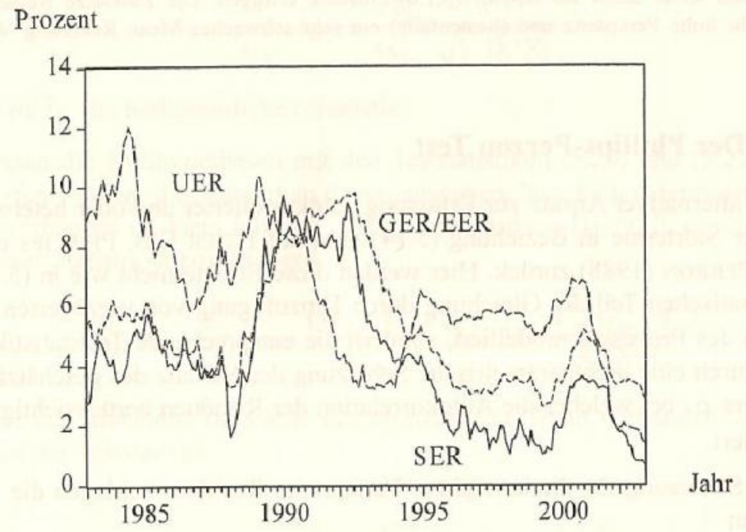


Abbildung 5.8: Entwicklung der schweizerischen, deutschen/europäischen und amerikanischen Euromarktsätze, Monatsdaten, Januar 1983 – Dezember 2002

Tabelle 5.2: Ergebnisse des Augmented Dickey-Fuller Tests
1/1983 – 12/2002, 240 Beobachtungen

Variable	Niveau k	Niveau Testwert	1. Differenz k	1. Differenz Testwert
SER	3	-1.194 (0.678)	2	-7.866 (0.000)
GER/EER	1	-0.957 (0.768)	0	-11.959 (0.000)
UER	1	-0.995 (0.755)	0	-11.151 (0.000)

Die Tests wurden für die Niveaus mit und für die ersten Differenzen ohne Konstante durchgeführt. Die Zahlen in Klammern geben die p-Werte an. Die Zahl der Lags, k, wurde mit Hilfe des Hannan-Quinn Kriteriums bestimmt.

Testing for Unit Roots

The unit root hypothesis corresponds to the null hypothesis. If we are unable to reject the null it does not necessarily mean that it is true. There could have been insufficient information to reject. Johnston/DiNardo fail to reject H_0 for simulated data from a stationary AR(1) with AR parameter 0.95: „Low power in statistical tests is often an unavoidable fact of life with which one must live and not expect to be able to make definitive pronouncements. Failure to reject a null hypothesis justifies at best only a cautious and provisional acceptance."

Alternative test by Kwiatkowski, Phillips, Schmidt and Shin (**KPSS test**).

Johnston, DiNardo (1997) "Econometric Methods, p. 227"; Verbeek (2000) "A Guide to Modern Econometrics "

Testing for Unit Roots

KPSS test

H_0: (trend-)stationary process

H_1: unit-root process

Idea:

1. Remove constant term or linear trend from series

   (to produce de-meaned or de-trended series)

2. Look at standardized, squared (partial) sums of residuals. These should diverge under H_1.

Kirchgässner, Wolters (2006), p. 161ff

Example: Long Memory in Inflation Rates

Augmented Dickey-Fuller Test and KPSS Tests for Inflation Rates
1/1969-9/1992, 285 Observations

	k	*U.S.*	*U.K.*	*France*	*Germany*	*Italy*
ADF Test	3	-4.43**	-4.48**	-2.71(*)	-4.98**	-3.31*
	6	-3.06*	-2.97*	-1.71	-3.49**	-2.24
	12	-1.86	-2.27	-1.29	-1.75	-2.39
KPSS Test	6	0.81**	1.02**	1.57**	1.26**	0.94**
	12	0.51*	0.65**	0.91**	0.80**	0.56**

(*), *, and **: respective null hypothesis can be rejected at 10%, 5% or 1% level

KPSS Test: reject H_0 → inflation rates are nonstationary

ADF: depends on k; for *k* = 3 stationary; for *k* = 6 unit root

Example: Long Memory in Inflation Rates

- → results are confusing ...
- ... maybe due to misspecification:
  ARIMA models with integer values of *d* (*d*=0 or *d*=1)
- Solution: **Fractionally integrated ARMA Models**

$$(1-L)^d y_t \quad \begin{cases} \text{stationary for } 0 < d < 0.5 \\ \text{nonstationary for } 0.5 < d < 1 \end{cases}$$

Hassler, Wolters (1995) “Long Memory in Inflation Rates: International Evidence”

Fractional Differencing

$$(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \ldots$$

$$= \sum_{j=0}^{\infty} d_j L^j \quad \text{with} \quad d_j = \binom{d}{j} = \frac{d \cdot (d-1) \cdot (d-(j-1))}{j!} \quad \text{and} \quad d_0 = 1$$

	d_0	d_1	d2	d3	d4
d=0.1	1	0.1	-0.045	0.0285	-0.02066
d=0.2	1	0.2	-0.080	0.0480	-0.03360
d=0.3	1	0.3	-0.105	0.0595	-0.04016
d=0.4	1	0.4	-0.120	0.0640	-0.04160
d=0.5	1	0.5	-0.125	0.0625	-0.03906

Hassler, Wolters (1995) “Long Memory in Inflation Rates: International Evidence”

Fractional Differencing

ARFIMA(0,*d*,0)

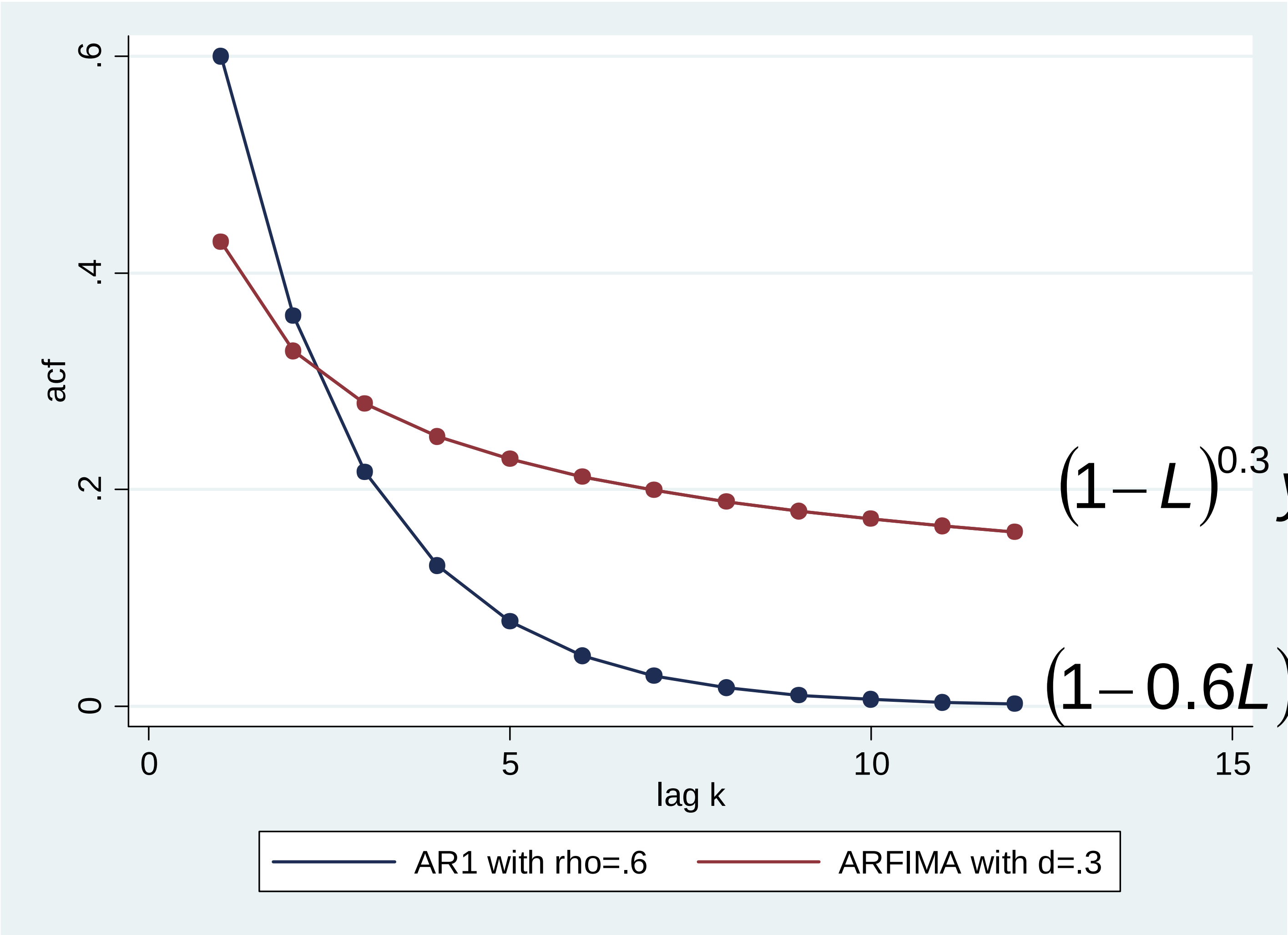
$$\left(1-L\right)^{d} y_t = \pounds_t$$

$$y_t = d\, y_{t-1} - \frac{1}{2} d(d-1)\, y_{t-2} + \frac{1}{6} d(d-1)(d-2)\, y_{t-3} - \ldots + \pounds_t$$

ARFIMA(p,*d*,q) $$a\left(L\right)\left(1-L\right)^{d} y_t = b\left(L\right)\varepsilon_t$$

Hassler, Wolters (1995) “Long Memory in Inflation Rates: International Evidence”

Fractional Differencing and Long Memory



$$(1-L)^{0.3} y_t = \varepsilon_t$$

$$(1-0.6L) y_t = \varepsilon_t$$

Multiple unit roots? $a(L)(1-L)^{d}y_t=\delta+b(L)\varepsilon_t$

- As a rule of thumb economic time series do **not** need to be differenced **more than two times**
- If you suspect two unit roots, estimate

$$\Delta^2 y_t=\theta\Delta y_{t-1}+\delta+\varepsilon_t$$

- If you can't reject H_0: $\theta=0$ $\rightarrow$ series has two unit roots
- If you reject, test for a single-unit root using

$$\Delta y_t=\delta+\pi_1 y_{t-1}+\pi_2\Delta y_{t-1}+\varepsilon_t$$

with H_0: $\pi_1=0$

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • **ARIMA Models** • Unit Root Tests • Seasonal ARIMA

ARIMA(*p*,*d*,*q*)

$$\underbrace{a(L)\underbrace{\underbrace{(1-L)^d y_t}_{\Delta^d y_t}}_{x_t}}\ = \delta + b(L)\varepsilon_t$$

$$\underbrace{x_t - \varphi_1 x_{t-1} - \ldots - \varphi_p x_{t-p}}_{a(L)x_t} = \delta + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}}_{b(L)\varepsilon_t}$$

with y_t ~ ARIMA(*p*,*d*,*q*) and x_t ~ ARMA(*p*,*q*)

Forecasting an ARIMA (*p*,1,*q*)

$$x_t = y_t - y_{t-1} \qquad \text{where } x_t \sim \text{ARMA}(p,q)$$

$$y_t = y_{t-1} + x_t$$

$$y_{T+1} = y_T + x_{T+1}$$

$$\rightarrow \tilde{y}_{T+1|T} = y_T + \tilde{x}_{T+1|T}$$

i.e. last observed level plus 1–step forecast of the change

$$y_{T+2} = y_{T+1} + x_{T+2} = y_T + x_{T+1} + x_{T+2}$$

$$\rightarrow \tilde{y}_{T+2|T} = y_T + \tilde{x}_{T+1|T} + \tilde{x}_{T+2|T}$$

Alternatively

$$\underbrace{a(L)(1-L)^d}_{A(L)} y_t = \delta + b(L)\varepsilon_t$$

$$A(L)y_t = \delta + b(L)\varepsilon_t$$

Solve for y_t (actually, solve for y_{T+l}) and use the "**general prediction formula**".

Example: ARIMA(1,1,0)

$$(1-0.74L)(y_t - y_{t-1}) = 0.021 + \varepsilon_t$$

$$y_t - 0.74y_{t-1} - y_{t-1} + 0.74y_{t-2} = 0.021 + \varepsilon_t$$

$$\underbrace{y_t - 174y_{t-1} + 0.74y_{t-2}}_{A(L)} = 0.021 + \varepsilon_t$$

$$y_t = 174y_{t-1} - 0.74y_{t-2} + 0.021 + \varepsilon_t$$

$$y_{T+l} = 174y_{T+l-1} - 0.74y_{T+l-2} + 0.021 + \varepsilon_{T+l}$$

Now use the "**general prediction formula**".

Example: ARIMA(1,1,0)

$$(1-\varphi_1 L)(1-L)y_t = \varepsilon_t$$

$$(1-\varphi_1 L)(y_t - y_{t-1}) = \varepsilon_t$$

$$y_t - \varphi_1 y_{t-1} - (y_{t-1} - \varphi_1 y_{t-2}) = \varepsilon_t$$

$$y_t = (1+\varphi_1)y_{t-1} - \varphi_1 y_{t-2} + \varepsilon_t$$

$$\tilde{y}_{T+1|T} = (1+\varphi_1)y_T - \varphi_1 y_{T-1}$$

$$\tilde{y}_{T+2|T} = (1+\varphi_1)\tilde{y}_{T+1|T} - \varphi_1 y_T$$

$$\tilde{y}_{T+l|T} = (1+\varphi_1)\tilde{y}_{T+l-1|T} - \varphi_1 \tilde{y}_{T+l-2|T}$$

Example: ARIMA(1,1,0) continued

$$\tilde{y}_{T+1|T} = (1+\varphi_1)y_T - \varphi_1 y_{T-1}$$

$$\tilde{y}_{T+2|T} = (1+\varphi_1)\tilde{y}_{T+1|T} - \varphi_1 y_T$$

$$\tilde{y}_{T+l|T} = (1+\varphi_1)\tilde{y}_{T+l-1|T} - \varphi_1 \tilde{y}_{T+l-2|T}$$

Repeatedly substituting yields:

$$\tilde{y}_{T+l|T} = y_T + (y_T - y_{T-1})\frac{\varphi_1(1-\varphi_1^l)}{(1-\varphi_1)}$$

As $l \to \infty$

$$\tilde{y}_{T+l|T} = y_T + (y_T - y_{T-1})\frac{\varphi_1}{(1-\varphi_1)}$$, i.e. a horizontal line

Example: ARIMA(0,1,1)

$$\Delta y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$y_t = y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Forecasts are constructed from the difference equation:

$$\tilde{y}_{T+l|T} = \tilde{y}_{T+l-1|T} + \underbrace{\tilde{\varepsilon}_{T+l|T}}_{=0} - \theta_1 \underbrace{\tilde{\varepsilon}_{T+l-1|T}}_{=0}, \quad l = 1,2,\ldots$$

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T$$

$$\tilde{y}_{T+l|T} = \tilde{y}_{T+1|T}, \quad l = 2,3,\ldots$$

Thus, for all lead time, the forecasts made at time *T* follow a horizontal straight line. (see example below)

Example: ARIMA(0,1,1)

The disturbance term, ε_T, is constructed from the difference equation:

$$\varepsilon_t = y_t - y_{t-1} + \theta_1 \varepsilon_{t-1} \quad t = 2,3,\ldots \quad \text{with } \varepsilon_1 = 0$$

Substituting repeatedly for lagged values gives:

$$\varepsilon_T = y_T + (1+\theta_1)\sum_{j=1}^{T-1}(-\theta_1)^j y_{T-j}$$

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T = (1+\theta_1)\sum_{j=0}^{T-1}(-\theta_1)^j y_{T-j}$$

ARIMA forecasting – ARIMA(1,1,0) vs ARIMA (0,1,1)

Compare:

depends on all past observations

ARIMA(0,1,1)

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T = (1+\theta_1)\sum_{j=0}^{T-1}(-\theta_1)^j y_{T-j}$$

ARIMA(1,1,0)

$$\tilde{y}_{T+l|T} = y_T + (y_T - y_{T-1})\frac{\varphi_1(1-\varphi_1^l)}{(1-\varphi_1)}$$

depends on last two observations

Example: ARIMA(0,1,1) continued

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T = (1+\theta_1)\sum_{j=0}^{T-1}(-\theta_1)^j y_{T-j}$$

$$= (1+\theta_1)\left[y_T + (-\theta_1)^1 y_{T-1} + (-\theta_1)^2 y_{T-2} + (-\theta_1)^3 y_{T-3} + \ldots\right]$$

Define $\lambda = (1+\theta_1) \quad \Rightarrow \quad -\theta_1 = 1-\lambda$

$$\tilde{y}_{T+1|T} = \lambda y_T + \lambda(1-\lambda)^1 y_{T-1} + \lambda(1-\lambda)^2 y_{T-2} + \lambda(1-\lambda)^3 y_{T-3} + \ldots$$

$$\tilde{y}_{T|T-1} = \lambda y_{T-1} + \lambda(1-\lambda)^1 y_{T-2} + \lambda(1-\lambda)^2 y_{T-3} + \lambda(1-\lambda)^3 y_{T-4} + \ldots$$

$$(1-\lambda)\tilde{y}_{T|T-1} = \lambda(1-\lambda) y_{T-1} + \lambda(1-\lambda)^2 y_{T-2} + \lambda(1-\lambda)^3 y_{T-3} + \ldots$$

$$\Rightarrow \tilde{y}_{T+1|T} = \lambda y_T + (1-\lambda)\tilde{y}_{T|T-1}$$

Exponential Smoothing (Exponentially Weighted Moving Average)

$$\tilde{y}_t = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots$$

$$\tilde{y}_{t-1} = \alpha y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \ldots$$

$$(1-\alpha)\tilde{y}_{t-1} = \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots$$

$$\tilde{y}_t = \alpha y_t + (1-\alpha)\tilde{y}_{t-1}$$

- The smaller α (where $0 \leq \alpha \leq 1$), the more heavily the smoothing
- Can be shown that $\alpha \sum_{\tau=0}^{\infty} (1-\alpha)^{\tau} = \frac{\alpha}{1-(1-\alpha)} = 1$
- For forecasting: $\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots$

  $\hat{y}_{T+1} = \alpha y_T + (1-\alpha)\hat{y}_T$

Example: ARIMA(0,1,1) continued

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T = (1+\theta_1)\sum_{j=0}^{T-1}(-\theta_1)^j y_{T-j}$$

This can be written as

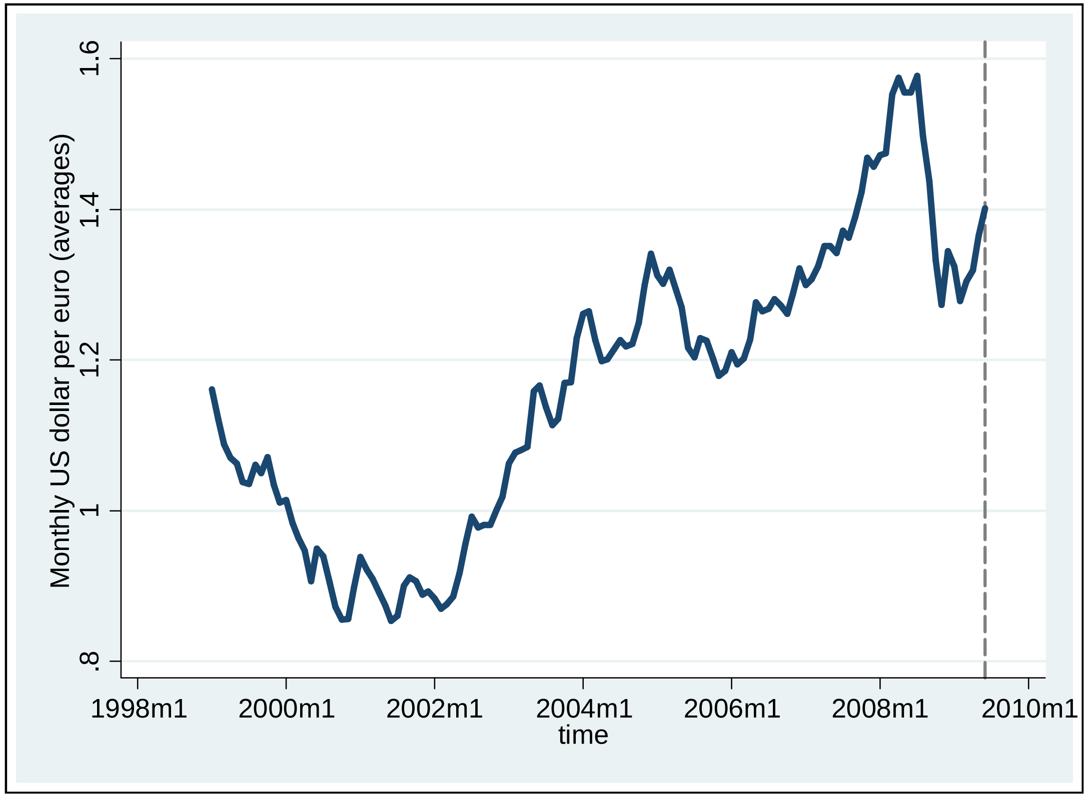
$$\tilde{y}_{T+1|T} = \lambda y_T + (1-\lambda)\tilde{y}_{T|T-1}$$

with $\lambda = (1+\theta_1)$

That is, ARIMA(0,1,1) is like an **EWMA**. The current forecast is a weighted average of the current observation and the previous forecast. However, this ARIMA version of EWMA estimates λ from the data and can provide prediction intervals

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Example: US dollar per euro (exchange rate)

No clear trend → **Case 2**: constant, no trend

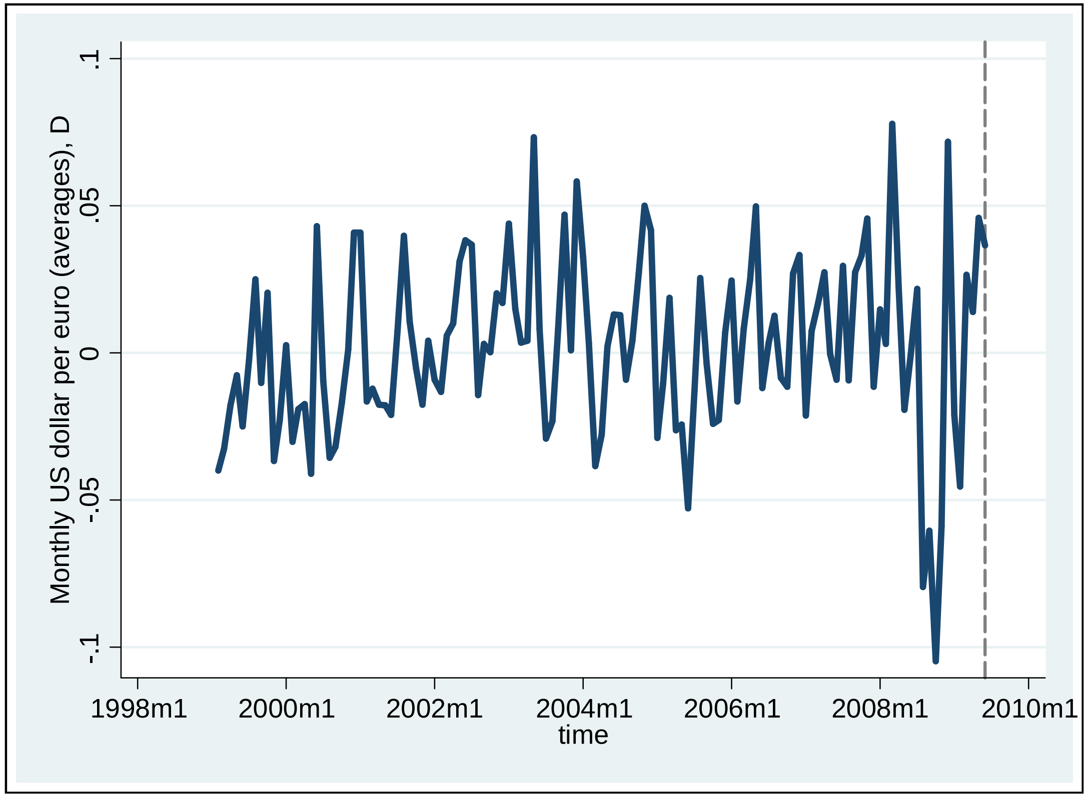
```
. dfuller exchange_rate if time <= 593,
Dickey-Fuller test for unit root                   Number of obs   =       125

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
                Statistic          Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.435            -3.502            -2.888            -2.578
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9041
```

DF	ADF(1)	ADF(2)	ADF(3)	ADF(4)	ADF(5)	ADF(6)
-0.435	-0.842	-0.712	-0.774	-0.810	-0.899	-0.757

→ the conclusion does not change, and **we cannot reject the presence of a first unit root**

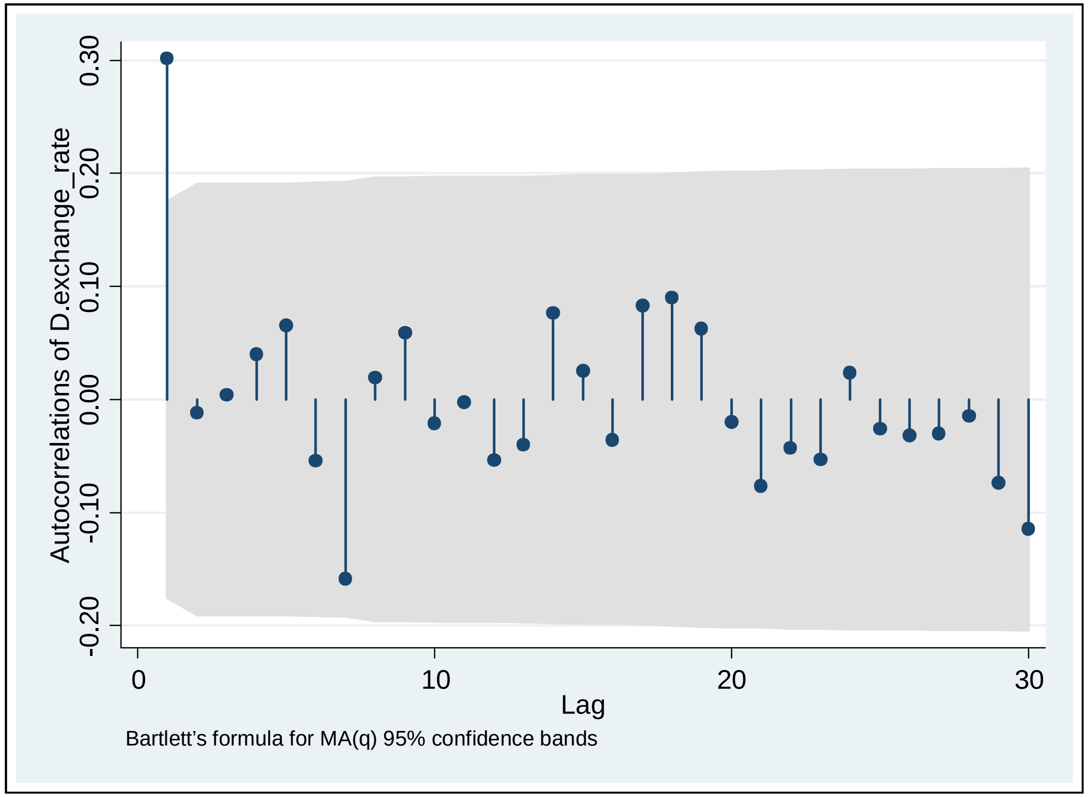
Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)



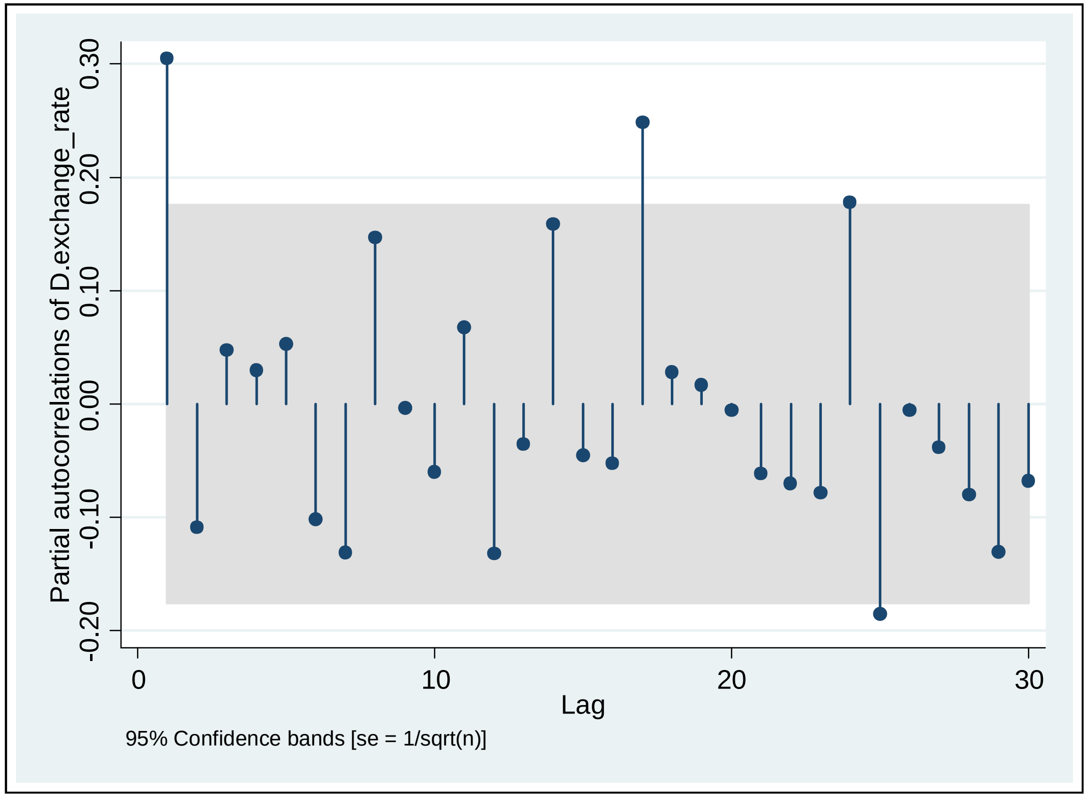
Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)



Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)



Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(0,1,1)
[...]
ARIMA regression

Sample:  1999m2 to 2009m6                        Number of obs      =       125
                                                 Wald chi2(1)       =     12.87
Log likelihood =  266.9772                       Prob > chi2        =    0.0003

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018907   .0034256     0.55   0.581    -.0048232    .0086047
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3452627    .096229     3.59   0.000     .1566573    .5338682
-------------+----------------------------------------------------------------
      /sigma |    .028576   .0015061    18.97   0.000     .0256241     .031528
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(1,1,0)
[...]
ARIMA regression

Sample:  1999m2 to 2009m6                    Number of obs      =       125
                                             Wald chi2(1)       =     17.43
Log likelihood =  266.1319                   Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018981    .003743     0.51   0.612     -.005438    .0092343
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3067845    .073477     4.18   0.000     .1627723    .4507968
-------------+----------------------------------------------------------------
      /sigma |   .0287709   .0015431    18.65   0.000     .0257465    .0317953
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(1,1,1)
[...]
ARIMA regression
Sample:  1999m2 to 2009m6                        Number of obs      =       125
                                                 Wald chi2(2)       =     12.87
Log likelihood =  266.9795                       Prob > chi2        =    0.0016
------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
exchange_r~e |
       _cons |   .0018907   .0034643     0.55   0.585    -.0048992    .0086806
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.0180361   .1774111    -0.10   0.919    -.3657554    .3296832
          ma |
         L1. |   .3611031   .1877277     1.92   0.054    -.0068364    .7290426
-------------+----------------------------------------------------------------
      /sigma |   .0285736   .0015466    18.48   0.000     .0255423    .0316049
------------------------------------------------------------------------------
```

Example: US dollar per euro (exchange rate)

Model	ARIMA(0,1,1)	ARIMA(1,1,0)	ARIMA(1,1,1)
AIC	-7.094	-7.081	-7.079
BIC	-7.033	-7.020	-6.956

$x_t \sim$ **ARMA(0,1)**

$$x_t = \varepsilon_t + 0.3471\varepsilon_{t-1}$$

$y_t \sim$ **ARIMA(0,1,1)**

$$(1-L)y_t = \varepsilon_t + 0.3471\varepsilon_{t-1}$$

Recall:

$$AIC = \log\left(\hat{\sigma}^2\right) + 2\frac{p+q}{T}$$

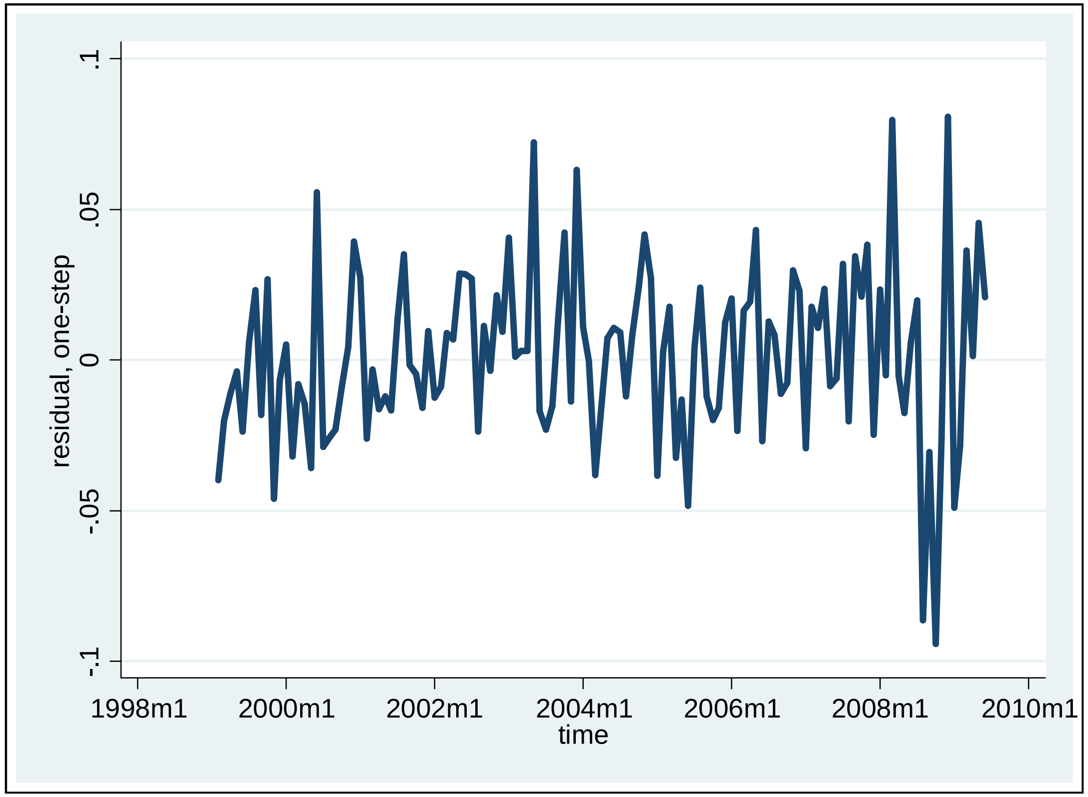
$$BIC = \log\left(\hat{\sigma}^2\right) + 2\frac{p+q}{T}\log T$$

Example: US dollar per euro (exchange rate)

```
. arima exchange_rate if time <=593, arima(0,1,1) noconst

[...]

ARIMA regression

Sample:  1999m2 to 2009m6                       Number of obs      =       125
                                                Wald chi2(1)       =     12.48
Log likelihood =  266.8257                      Prob > chi2        =    0.0004

------------------------------------------------------------------------------
             |                 OPG
D.           |
exchange_r~e |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .3470594   .0982231     3.53   0.000     .1545457    .5395731
-------------+----------------------------------------------------------------
      /sigma |   .0286087   .0015099    18.95   0.000     .0256493     .031568
------------------------------------------------------------------------------
```

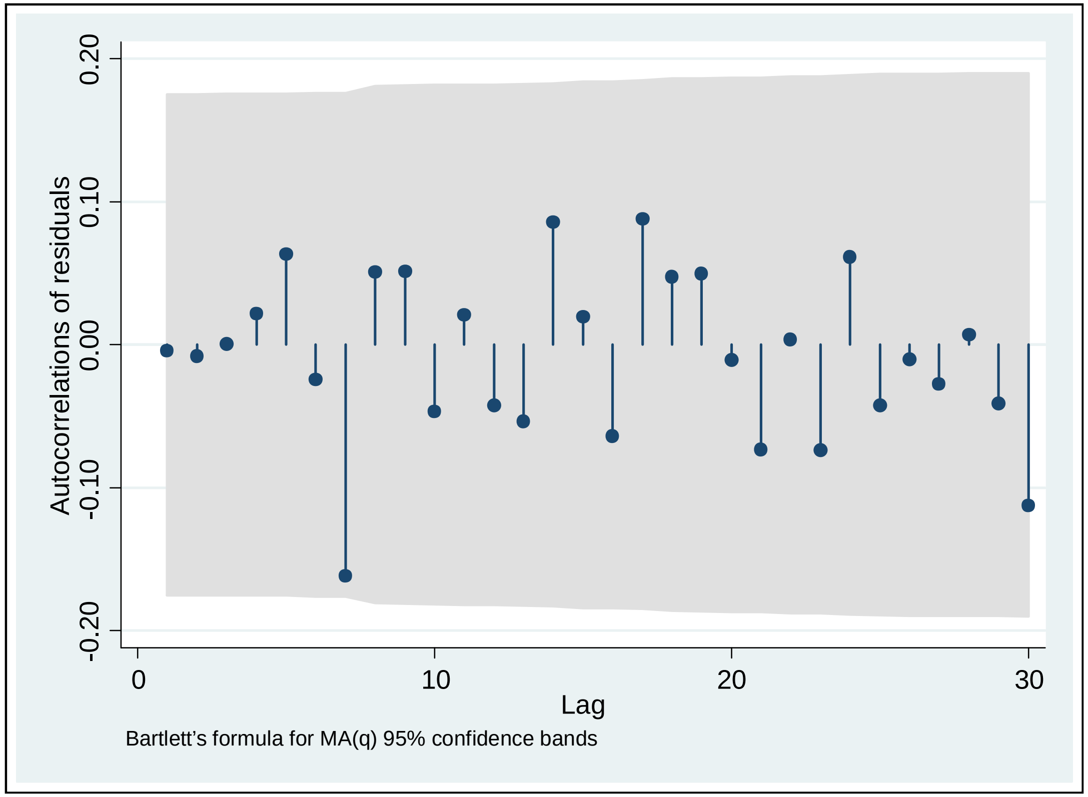
Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)



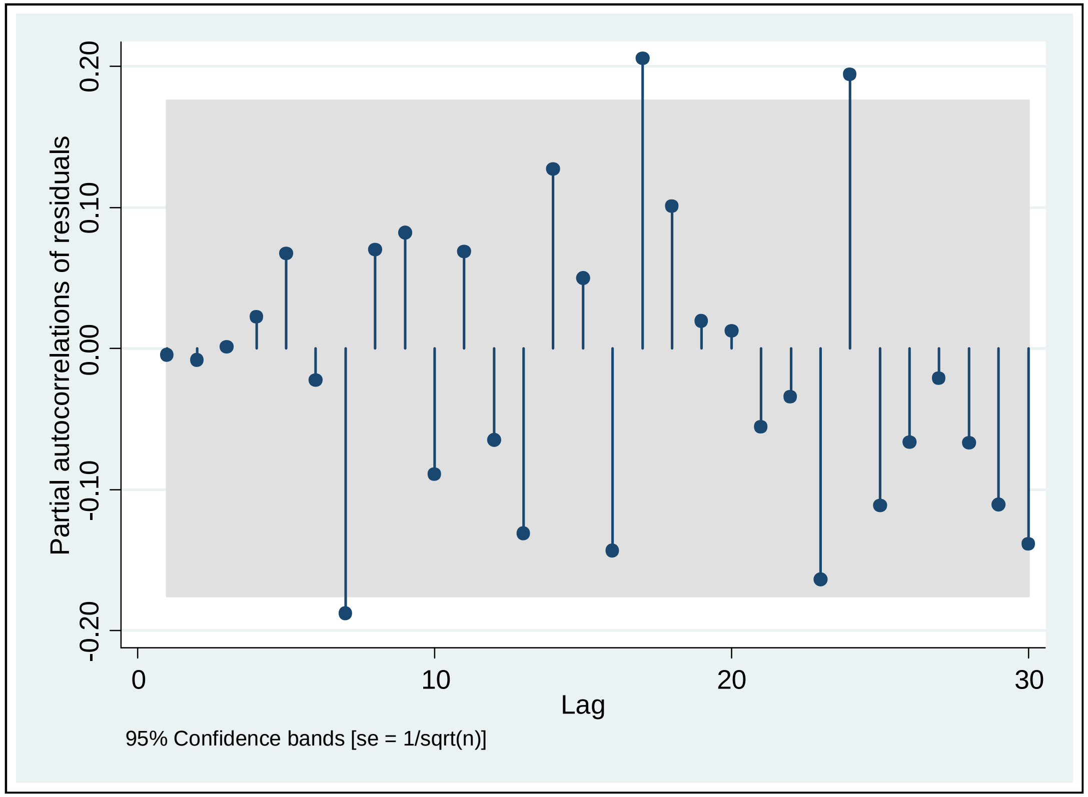
Example: US dollar per euro – **First Difference**

January 1999 to June 2009 (December 2009)



Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)



Example: US dollar per euro – First Difference

January 1999 to June 2009 (December 2009)

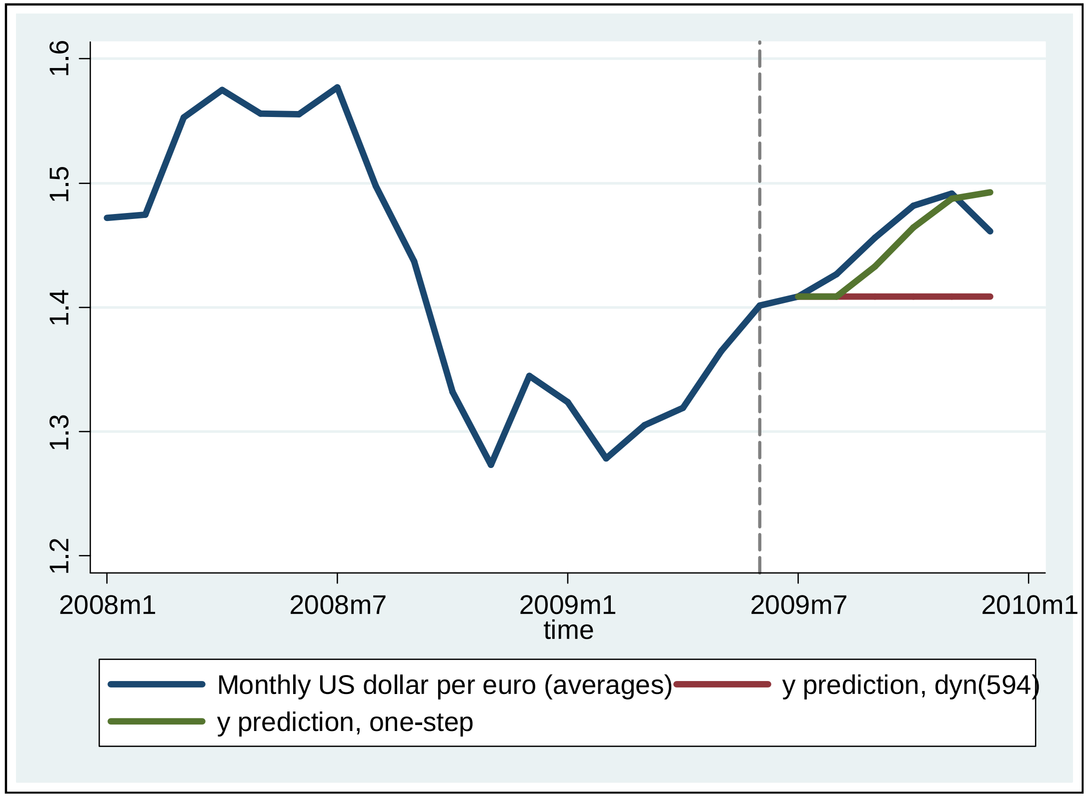
Recall:

$$Q = T(T+2)\sum_{k=1}^{K} \frac{1}{T-K}\hat{\rho}_k^2$$

```
. corrgram residuals, lags(30)
                                                  -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        -0.0044  -0.0044   .0025  0.9601          |                   |
2        -0.0080  -0.0081  .01085  0.9946          |                   |
3         0.0008   0.0009  .01093  0.9997          |                   |
4         0.0214   0.0224  .07111  0.9994          |                   |

[...]

19        0.0500   0.0196  9.4702  0.9648          |                   |
20       -0.0105   0.0125  9.4867  0.9766          |                   |
21       -0.0731  -0.0556  10.302  0.9747          |                   |
22        0.0038  -0.0346  10.304  0.9833          |                   |
23       -0.0738  -0.1635  11.152  0.9816          |                  -|
24        0.0612   0.1939  11.741  0.9827          |                   |-
25       -0.0422  -0.1114  12.023  0.9864          |                   |
26       -0.0102  -0.0666   12.04  0.9909          |                   |
27       -0.0275  -0.0214  12.162  0.9936          |                   |
28        0.0068  -0.0671  12.169  0.9959          |                   |
29       -0.0408  -0.1104  12.445  0.9969          |                   |
30       -0.1128  -0.1389  14.572  0.9919          |                  -|
```

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Recall

$$\tilde{y}_{T+1|T} = y_T - \theta_1 \varepsilon_T$$

$$\tilde{y}_{T+l|T} = \tilde{y}_{T+1|T}, \quad l = 2,3,\ldots$$

$$\tilde{y}_{T+1|T} = \lambda y_T + (1-\lambda)\tilde{y}_{T|T-1}$$

with $\lambda = (1+\theta_1)$

Here

$$y_t = y_{t-1} + \varepsilon_t + 0.3471\varepsilon_{t-1}$$

and $\hat{\theta} =$ -0.3471

$\hat{\lambda} = 0.6529$

For any unit root model

$$(1-L)y_t = \delta + \psi(L)\varepsilon_t$$

$$\Delta y_t = \delta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-1} + \ldots$$

Here: $(1-L)y_t = \varepsilon_t + 0.3471\varepsilon_{t-1}$

Hence: $a(L)c(L) = b(L)$

$$1\cdot\left(1+\psi_1 L + \psi_2 L^2 + \ldots\right) = \left(1+0.3471L\right)$$

$$\Rightarrow \psi_1 = 0.3471, \quad \psi_2 = \psi_3 = \ldots = 0$$

Recall for unit root models

The *s*-period-ahead forecast error

$$y_{t+s|t} - \hat{y}_{t+s|t} = \varepsilon_{t+s} + \{1+\psi_1\}\varepsilon_{t+s-1} + \{1+\psi_1+\psi_2\}\varepsilon_{t+s-2} + \dots$$
$$+ \{1+\psi_1+\psi_2+\dots+\psi_{s-1}\}\varepsilon_{t+1}$$

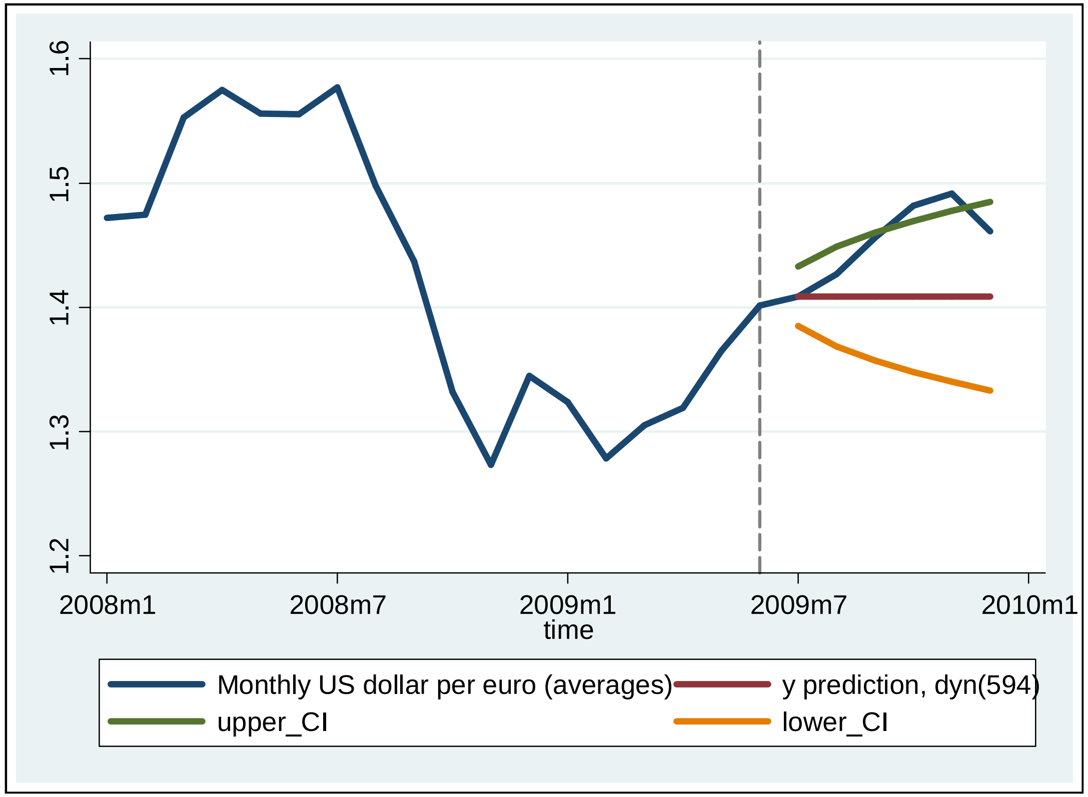
$$E\left(y_{t+s|t} - \hat{y}_{t+s|t}\right)^2 = \{1+(1+\psi_1)^2+(1+\psi_1+\psi_2)^2+\dots$$
$$+(1+\psi_1+\psi_2+\dots+\psi_{s-1})^2\}\sigma_\varepsilon^2$$

This can be used here but simplifies greatly because all ψs, except for ψ_1, are 0.

Hamilton (1994) “Time Series Analysis”, p. 435-442

Example: US dollar per euro (exchange rate)

January 1999 to June 2009 (December 2009)



Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • **Seasonal ARIMA**

Recall: ARIMA

ARIMA(*p*,*d*,*q*)

$$a(L)\underbrace{\underbrace{(1-L)^d y_t}_{\Delta^d y_t}}_{x_t} = \delta + b(L)\varepsilon_t$$

$$\underbrace{x_t - \varphi_1 x_{t-1} - \ldots - \varphi_p x_{t-p}}_{a(L)x_t} = \delta + \underbrace{\varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}}_{b(L)\varepsilon_t}$$

with $\boldsymbol{y_t}$ **~ ARIMA(*p*,*d*,*q*)** and $\boldsymbol{x_t}$ **~ ARMA(*p*,*q*)**

Seasonality

When observations are available on a monthly (s = 12) or a quarterly (s = 4) basis, some allowance must be made for seasonal effects. Two approaches:

- Work with seasonally adjusted data
- Incorporate seasonality into time series models

Multiplicative Seasonal ARIMA

- allows for stationary seasonal pattern that tends to disappear with increasing lag or lead time
- allows also for nonstationary seasonal pattern ("seasonal trend", slowly changing seasonal pattern)

Harvey (1981) "Time Series Models", p. 171-185

Seasonal Adjustment $y_t = L_t \cdot C_t \cdot S_t \cdot I_t \quad t = 1,\dots,T$

The objective is to eliminate the seasonal component *S*:

1. Isolate the combined long-term trend and cyclical components by removing the combined seasonal and irregular components.
2. Divide the original data by the smoothed series to estimate the combined seasonal and irregular components:
$$\frac{y_t}{\tilde{y}_t} = \frac{L_t \cdot C_t \cdot S_t \cdot I_t}{L_t \cdot C_t} = S_t \cdot I_t$$
3. Eliminate the irregular component as completely as possible. Average the values of the combined seasonal and irregular components corresponding to the same period. These averages will then be estimates of the seasonal indices.
4. Deseasonalize the original series by dividing each value by its corresponding seasonal index.

Why incorporate seasonality in the model?

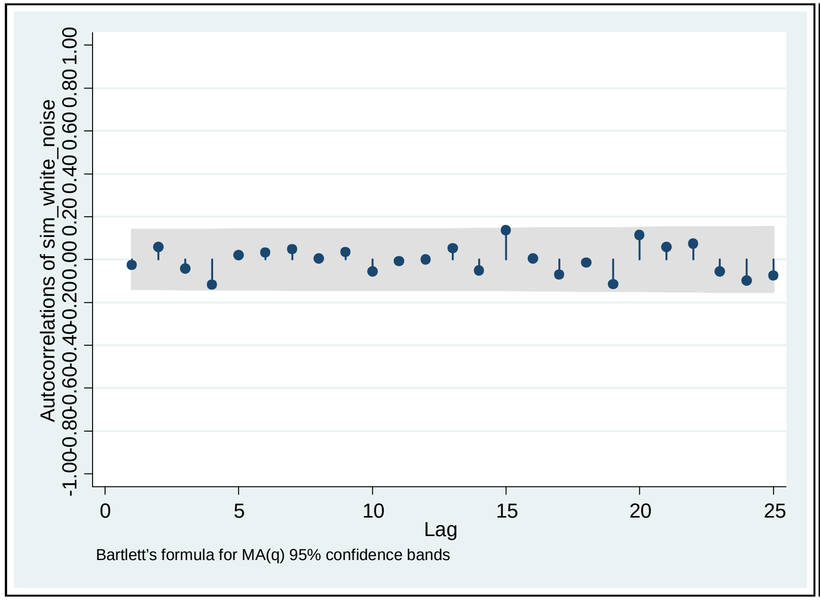
Seasonal adjustment procedures tend to result in over-adjustment, so that there is a tendency for seasonally adjusted series to exhibit negative autocorrelations at the seasonal lags.

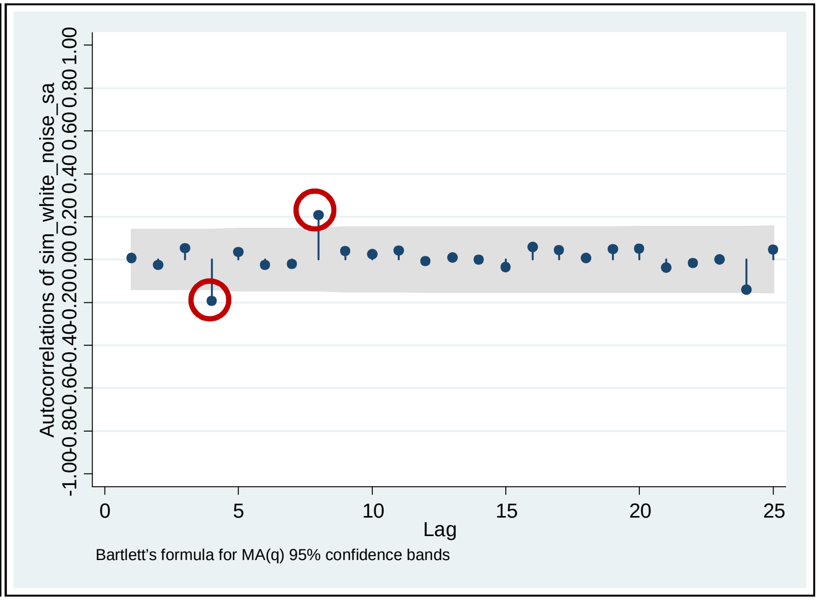
It is **generally preferable to work with the unadjusted series**. Seasonal adjustment can introduce considerable distortion into a series, and, at the same time, there is no guarantee that the adjusted series will be free from seasonal effects.

Harvey (1981) “Time Series Models”, p. 171-185

Why incorporate seasonality in the model?

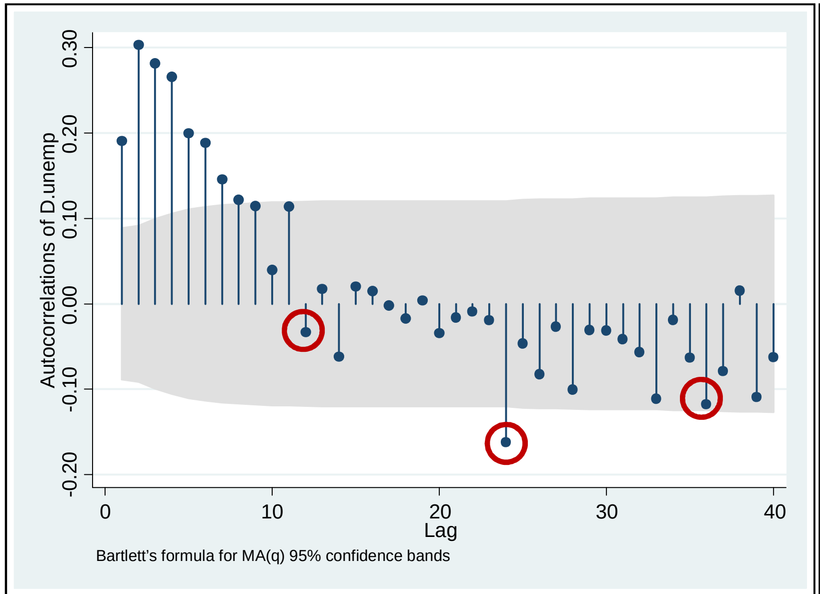
ACF of a **simulated white noise process** with 200 observations	ACF of a **seasonally adjusted (using a procedure for quarterly data) simulated white noise** process with 200 observations

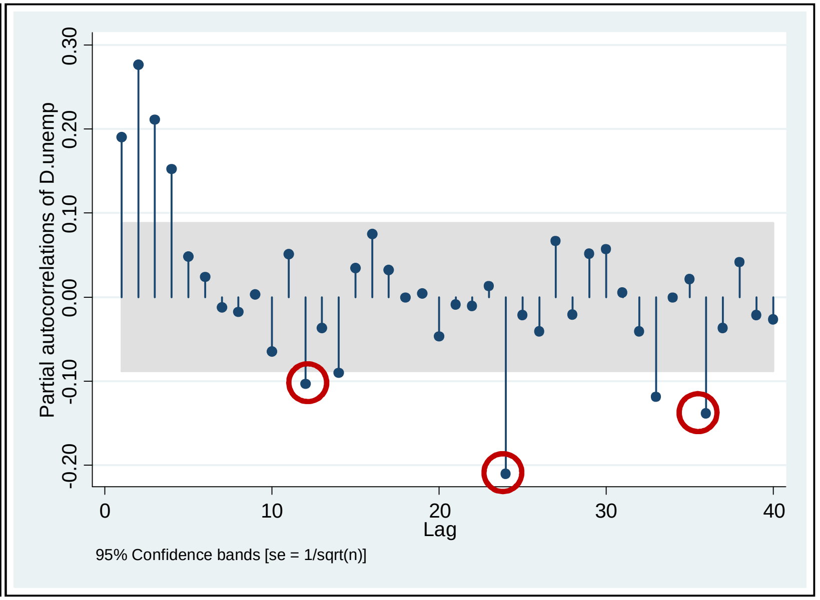




Why incorporate seasonality in the model?

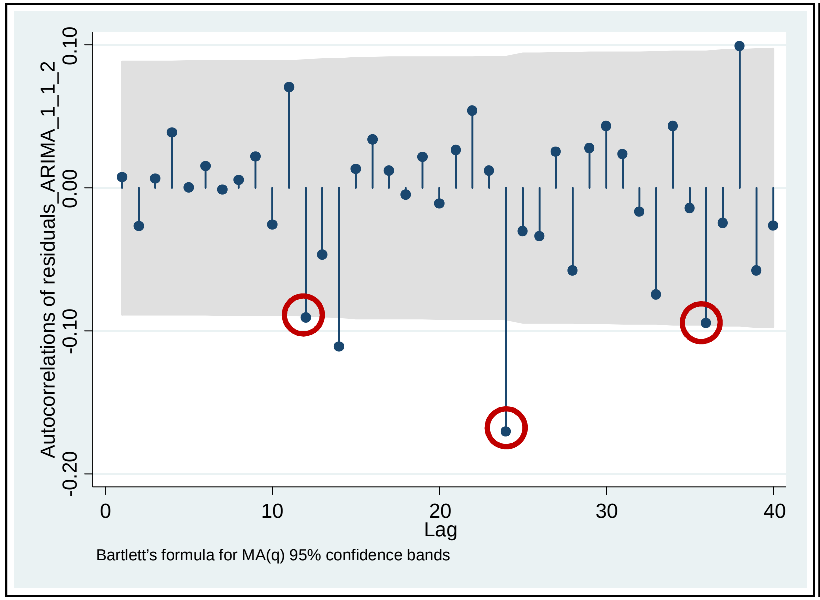
ACF of the first differenced monthly U.S. Unemployment Rate (1969m1 to 2009m12), **seasonal adjusted**	PACF of the first differenced monthly U.S. Unemployment Rate (1969m1 to 2009m12), **seasonal adjusted**

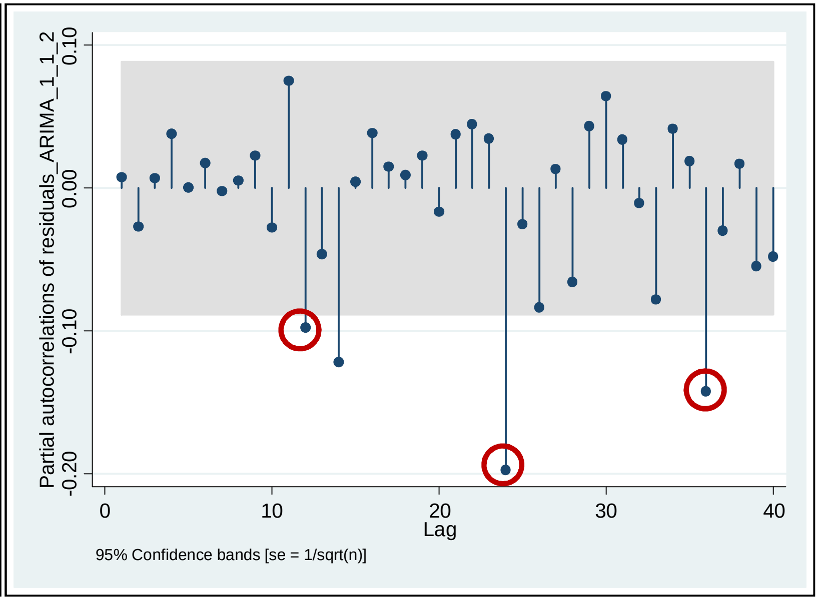




Why incorporate seasonality in the model?

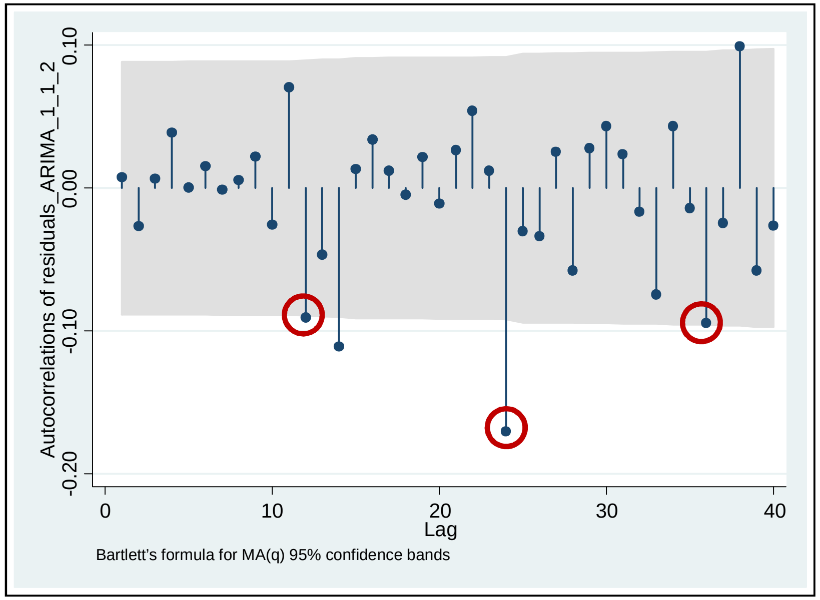
ACF of the residuals of an ARIMA(1,1,2) model fitted to the monthly U.S. Unemployment Rate (1969m1 to 2009m12), **seasonal adjusted**	PACF of the residuals of an ARIMA(1,1,2) model fitted to the monthly U.S. Unemployment Rate (1969m1 to 2009m12), **seasonal adjusted**





Why incorporate seasonality in the model?

ACF of the residuals of an ARIMA(1,1,2) model fitted to the monthly U.S. Unemployment Rate (1969m1 to 2009m12),
seasonal adjusted




```
. corrgram residuals_ARIMA_1_1_2

LAG        AC        PAC        Q      Prob>Q
-----------------------------------------
10       -0.0257   -0.0279    1.851   0.9974
11        0.0707    0.0748   4.3719   0.9578
12       -0.0904   -0.0975   8.5006   0.7449
13       -0.0466   -0.0461   9.6008   0.7262
[...]
22        0.0542    0.0448   18.801   0.6576
23        0.0121    0.0347   18.877   0.7083
24       -0.1702   -0.1971   33.901   0.0865
25       -0.0301   -0.0256    34.37   0.1002
[...]
35       -0.0138    0.0190   42.904   0.1685
36       -0.0947   -0.1421    47.67   0.0923
```

Types of Seasonal Effects

Permanent

- repeated more or less regularly (“periodic”): “every January”…
- may be regarded as stemming from factors such as the weather
- modeling such effects is quite similar to modeling trends

Temporary

- stationary model, in which any seasonal pattern tends to disappear
- Example: A dock strike in March of last year could influence production targets in March of this year (correlation of observations in the same month of different years), if firms believe that there is a high probability of such an event happening again. However, unless dock strikes in March turn out to be a regular occurrence, the effect of the original strike will be transitory (correlations can be expected to be small if the years are a long way apart).

Harvey (1981) “Time Series Models”, p. 171-185

Example: Purely Seasonal ARIMA(0,1,1)$_S$ Process

Typically, $s = 12$ or $s = 4$

$$(1 - L^s) y_t = (1 - \theta_1^s L^s) \zeta_t$$

with $\zeta_t \sim i.i.d., E(\zeta_t) = 0$ and $Var(\zeta_t) = \pm_\zeta^2$

Regular, periodic seasonality

Transitory seasonality

$$(1 - L^s) y_t = (1 - \theta_1^s L^s) \zeta_t$$

$$y_t - y_{t-s} = \zeta_t - \theta_1^s \zeta_{t-s}$$

$$y_t = y_{t-s} + \zeta_t - \theta_1^s \zeta_{t-s}$$

Harvey (1981) “Time Series Models”, p. 171-185

Example: Purely Seasonal ARIMA $(0,1,1)_S$ Process

$$(1-L^s)y_t = (1-\theta_1^s L^s)\zeta_t$$

$$y_t - y_{t-s} = \zeta_t - \theta_1^s \zeta_{t-s}$$

$$y_t = y_{t-s} + \zeta_t - \theta_1^s \zeta_{t-s}$$

In General: Purely Seasonal ARIMA $(P,D,Q)_S$

$$(1-\Phi_1^s L^s - \ldots - \Phi_P^s L^{P_s})(1-L^s)^D y_t = (1-\theta_1^s L^s - \ldots - \theta_Q^s L^{Q_s})\zeta_t$$

$$\Phi^s(L^s)\Delta_s^D y_t = \theta^s(L^s)\zeta_t$$

Harvey (1981) "Time Series Models", p. 171-185

Overview

Purely seasonal ARMA of order $(P,Q)_S$

$$\ddot{\ }^{s}\left(L^{s}\right)y_{t}=\theta^{s}\left(L^{s}\right)\zeta_{t}$$

Purely seasonal ARIMA of order $(P,D,Q)_S$

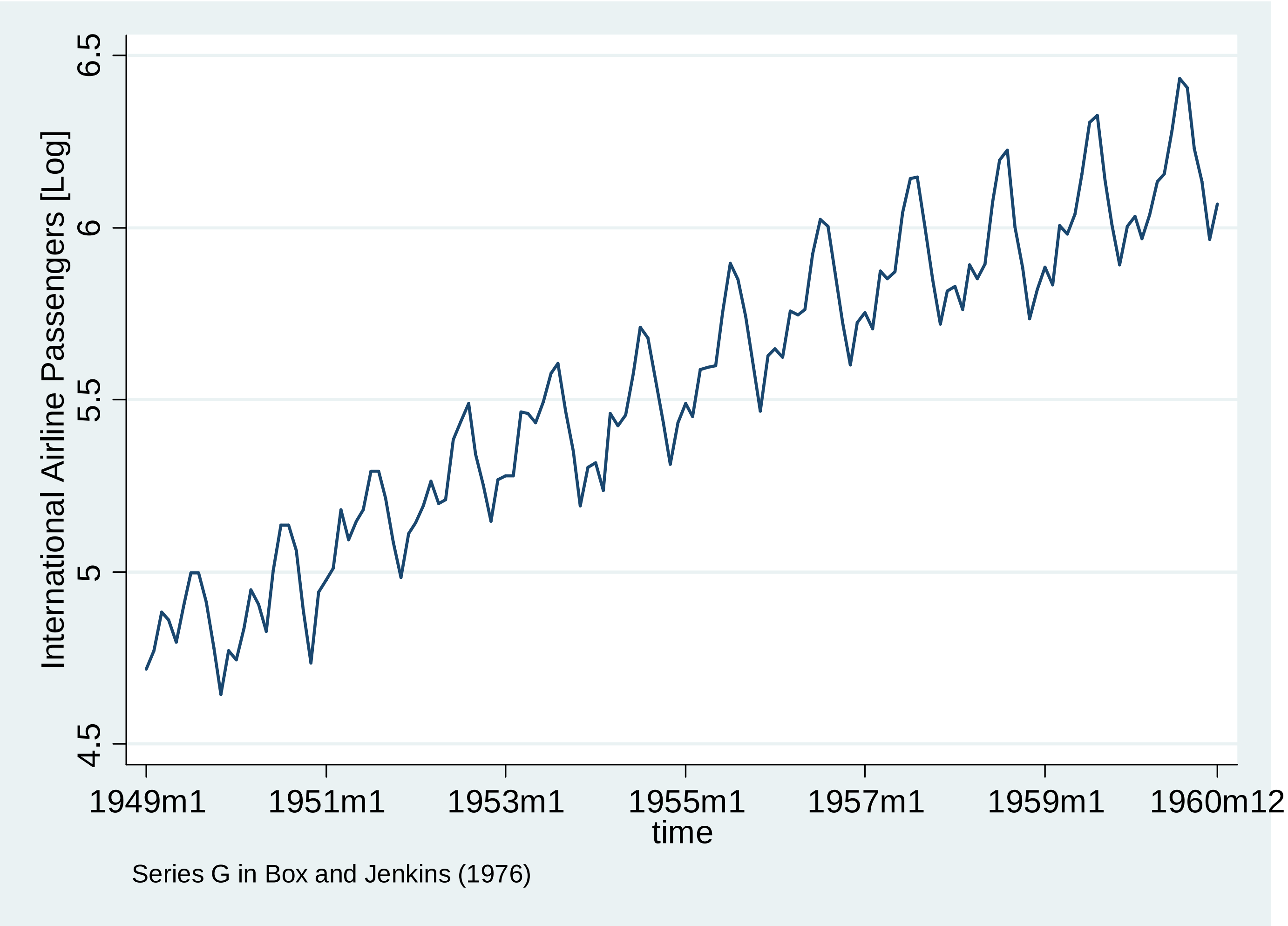
$$\ddot{\ }^{s}\left(L^{s}\right)\Delta_{s}^{D}y_{t}=\theta^{s}\left(L^{s}\right)\zeta_{t}$$

Multiplicative seasonal ARIMA
of order $(p,d,q)\times(P,D,Q)_S$

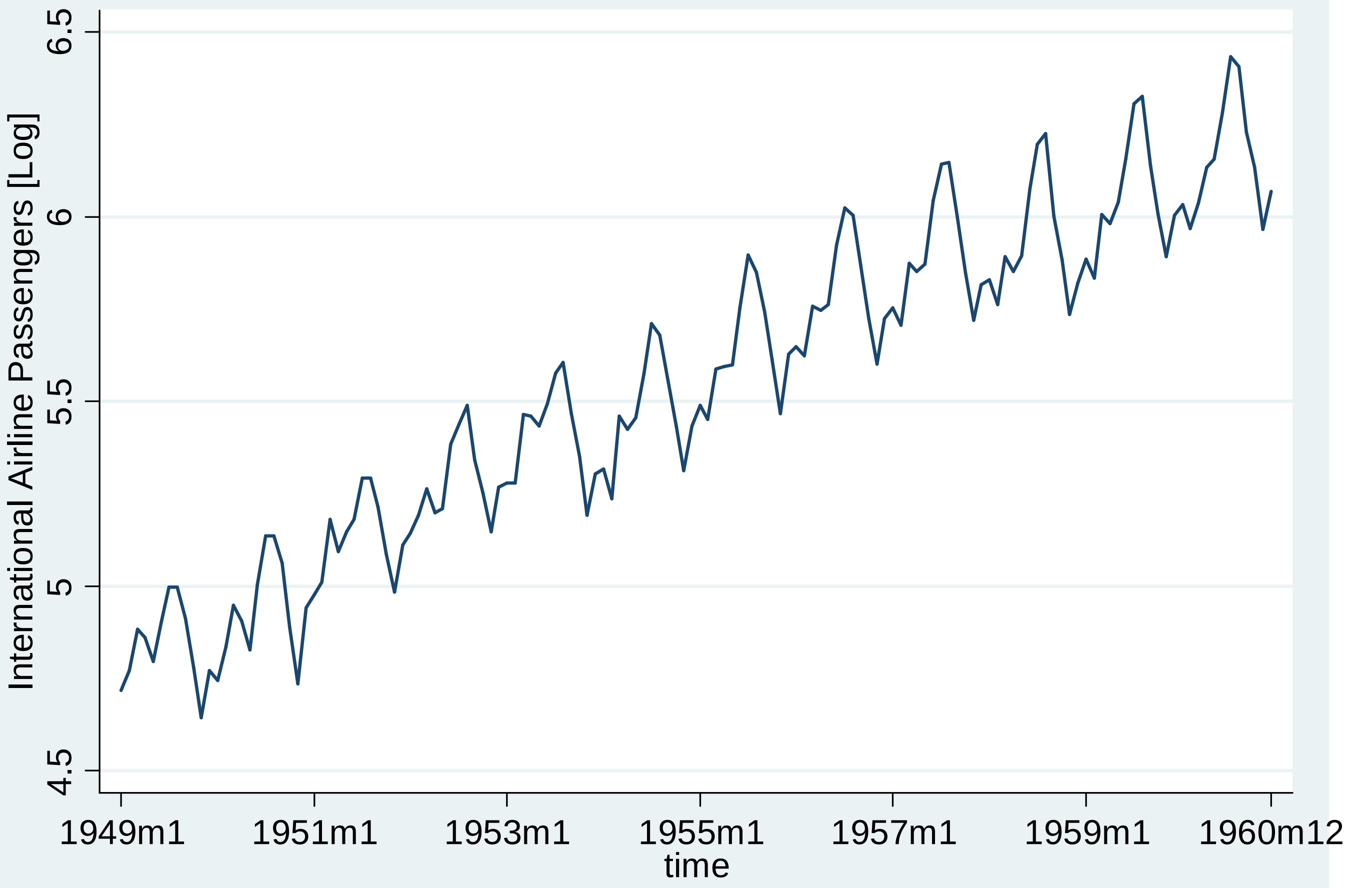
$$\ddot{\ }^{s}\left(L^{s}\right)\ddot{\ }\left(L\right)\Delta^{d}\Delta_{s}^{D}y_{t}=\theta^{s}\left(L^{s}\right)\theta\left(L\right)\varepsilon_{t}$$

Harvey (1981) "Time Series Models", p. 171-185

Preview: **Multiplicative seasonal ARIMA ?**



Preview: **Multiplicative seasonal ARIMA ?**



Series G in Box and Jenkins (1976)

$$\text{ARIMA}\,(0,1,1)\times(0,1,1)_{12}$$

Primary distinguishing characteristics of theoretical ACF's and PACF's for stationary processes

Process	ACF	PACF
AR	Tails off toward zero (exponential decay or damped sine wave)	Cuts off to zero (after lag *p*)
MA	Cuts off to zero (after lag *q*)	Tails off toward zero (exponential decay or damped sine wave)
ARMA	Tails off toward zero	Tails off toward zero

Pankratz (1983), Forecasting with univariate Box-Jenkins models, p.122

Purely Seasonal ARMA Processes

Purely Seasonal AR(1)

Stationary series of quarterly observations (*s* = 4)

$$\left(1-\Phi_1^s L^s\right) y_t = \zeta_t$$

Can be viewed as an AR(4) process with constraint

$$y_t = \Phi_4 y_{t-4} + \varepsilon_t \qquad \left|\Phi_4\right| < 1$$

$$\Phi_1 = \Phi_2 = \Phi_3 = 0$$

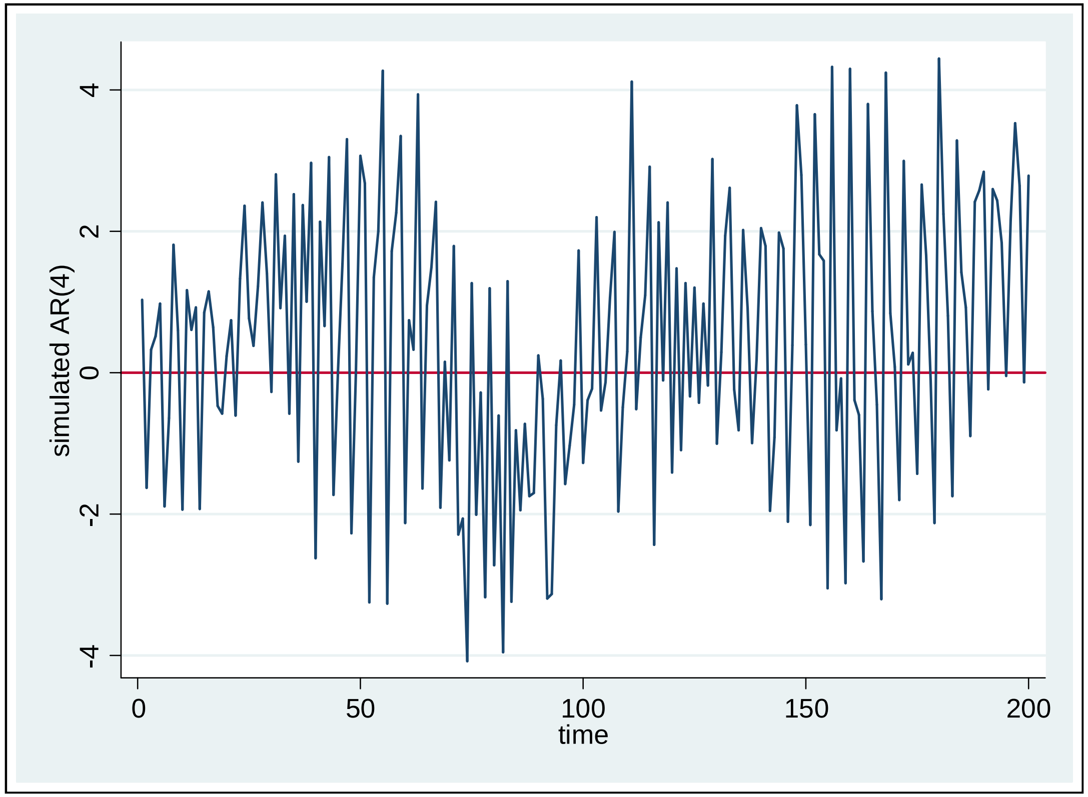
Autocorrelation function: $\rho_k = \begin{cases} \Phi_4^{k/4} & k = 0,4,8,\ldots \\ 0 & \text{otherwise} \end{cases}$

The closer $\left|\Phi_4\right|$ is to unity, the stronger the seasonal pattern.

Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal AR(1)

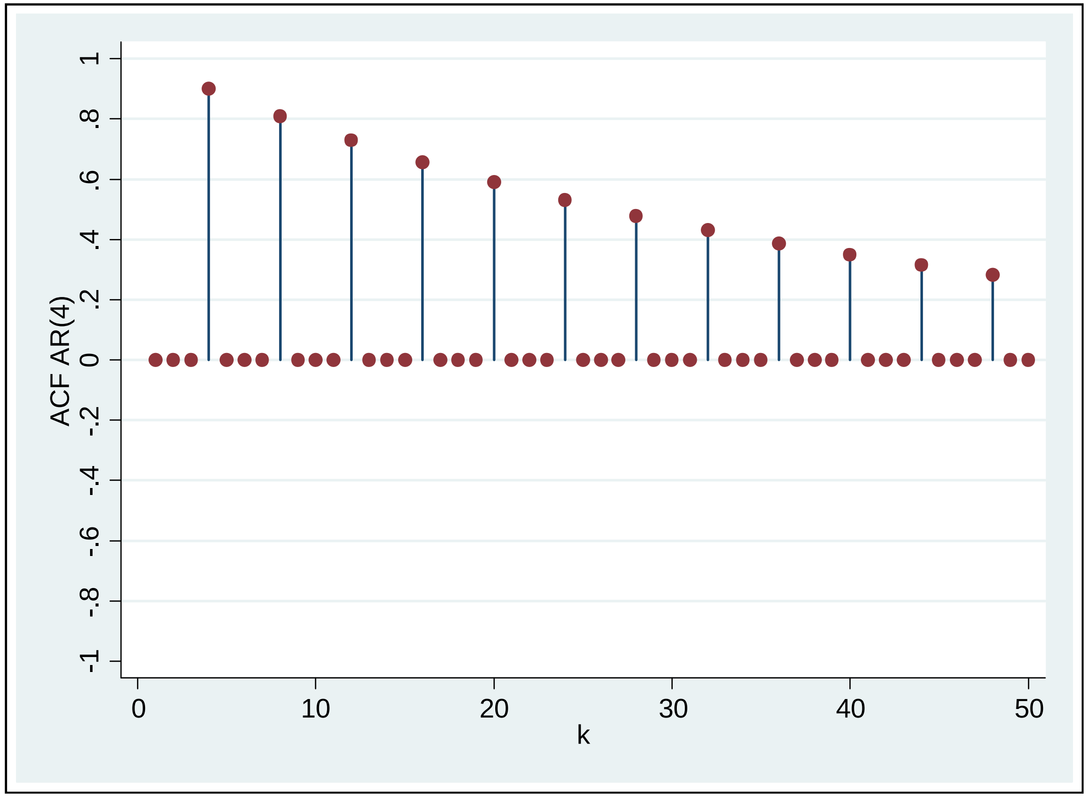
Simulated AR(1) with s = 4 and φ_4 = 0.9



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal AR(1)

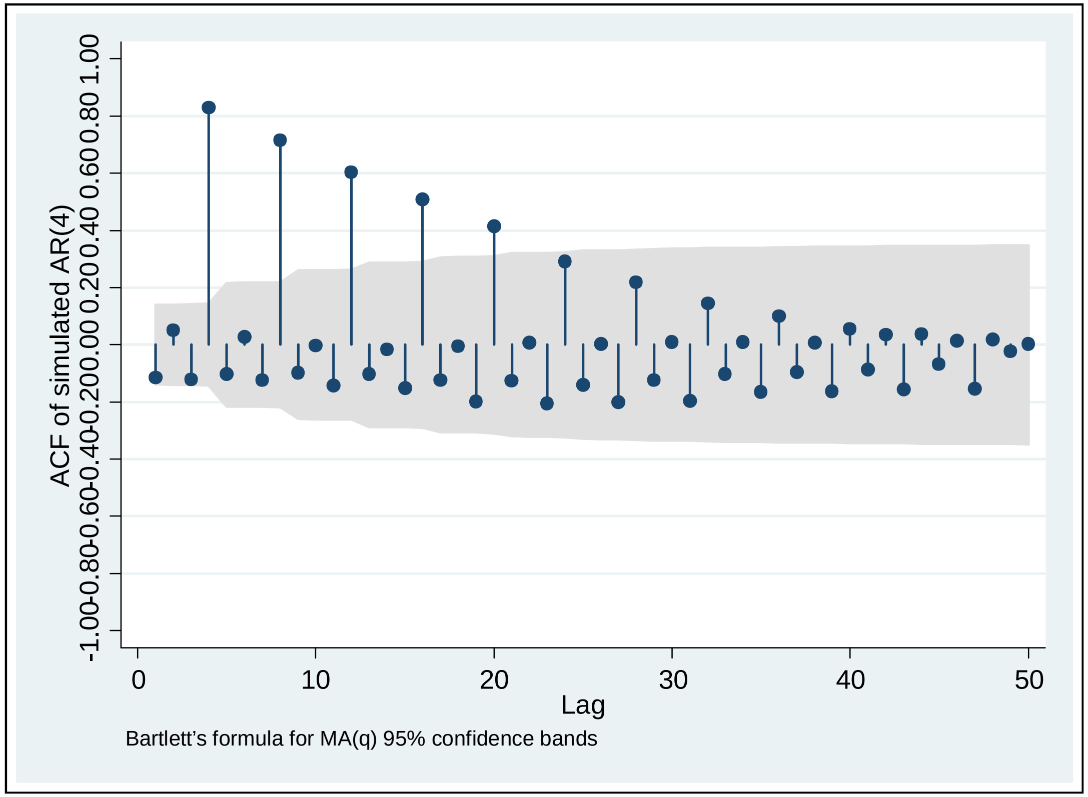
Theoretical ACF of an AR(1) with s = 4 and $\varphi_4 = 0.9$



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal AR(1)

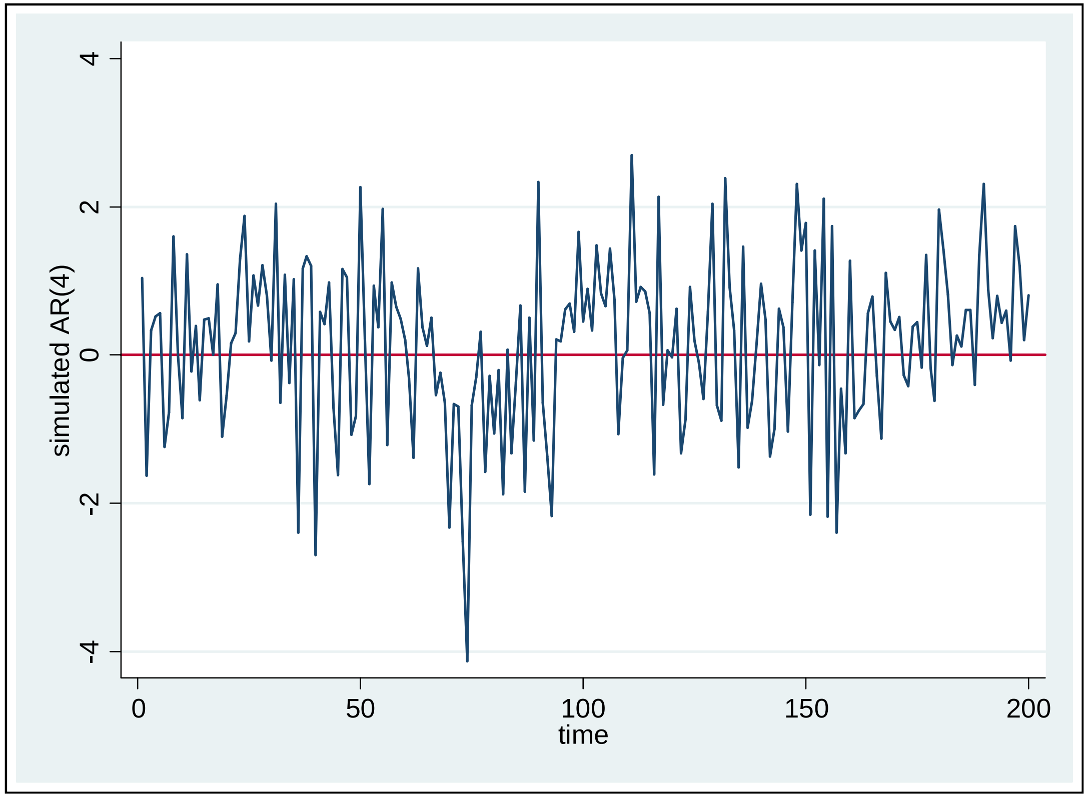
ACF of an simulated AR(1) with $s = 4$ and $\varphi_4 = 0.9$



Harvey (1981) "Time Series Models", p. 171-185

Purely Seasonal AR(1)

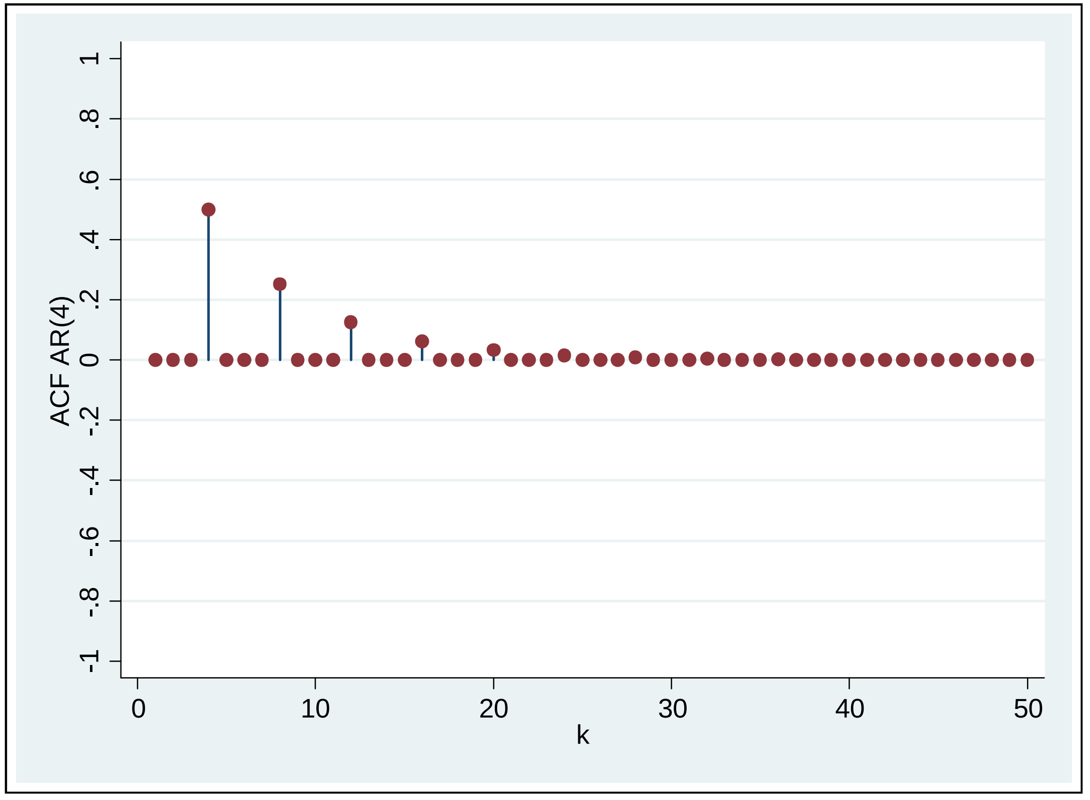
Simulated AR(1) with s = 4 and $\varphi_4 = 0.5$



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal AR(1)

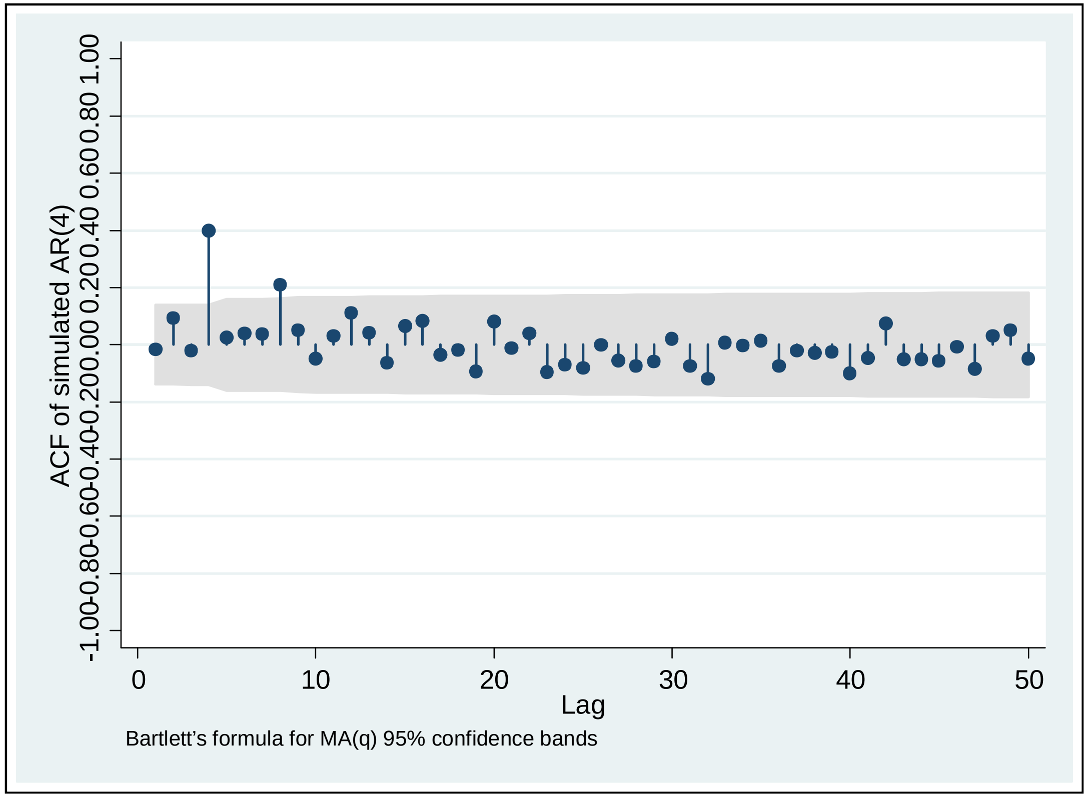
Theoretical ACF of an AR(4) with $s = 4$ and $\varphi_4 = 0.5$



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal AR(1)

ACF of an simulated AR(1) with $s = 4$ and $\varphi_4 = 0.5$



Harvey (1981) "Time Series Models", p. 171-185

ACF of Purely Seasonal AR(1)

$$\left(1-\text{¨}_1^{\,s} L^s\right) y_t = \zeta_t$$

Recall: Autocorrelation function for an AR(1) process

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k$$

$$\text{for } s = 4\text{:} \qquad \rho_k = \begin{cases} \left(\text{¨}_1^{\,s}\right)^{k/4} & k = 0,4,8,\ldots \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } s = 12\text{:} \qquad \rho_k = \begin{cases} \left(\text{¨}_1^{\,s}\right)^{k/12} & k = 0,12,24,\ldots \\ 0 & \text{otherwise} \end{cases}$$

Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal MA(1)

Stationary Series of monthly observations (*s* = 12)

$$y_t = \left(1-\theta_1^s L^s\right)\zeta_t$$

Can be viewed as an MA(12) process with constraint

$$y_t = \varepsilon_t - \theta_{12}\varepsilon_{t-12}$$

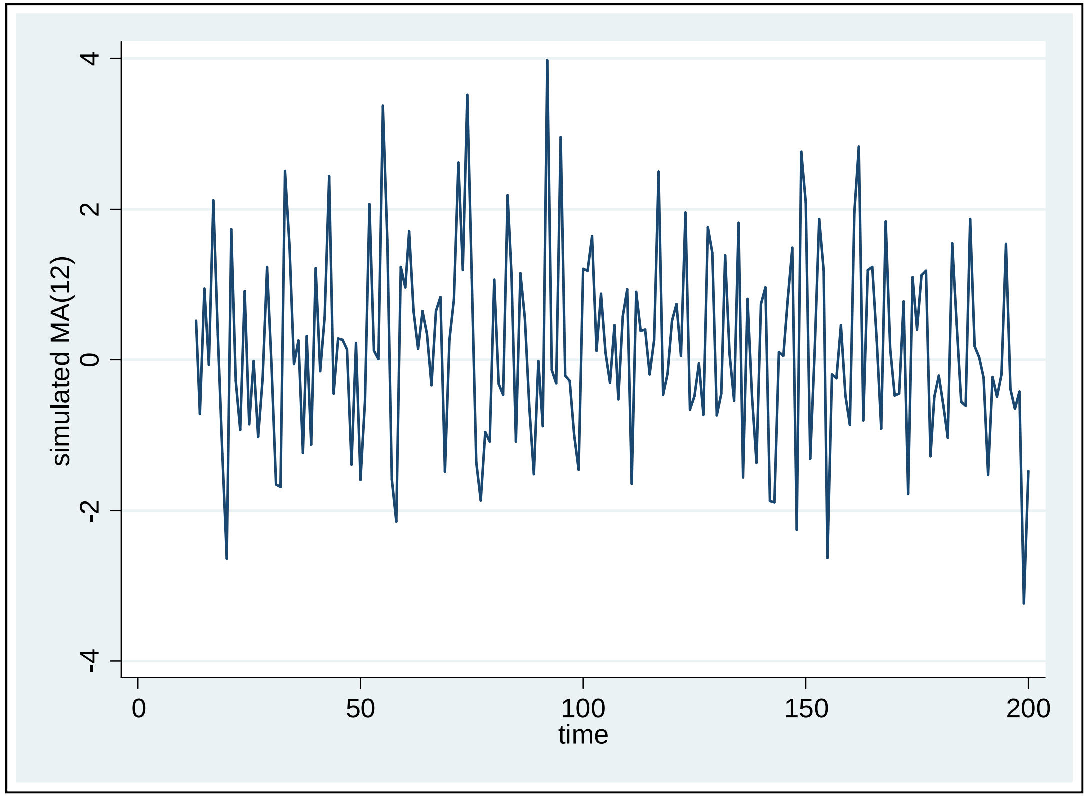
$$\theta_1 = \theta_2 = \theta_3 = \ldots = \theta_{11} = 0$$

Autocorrelation function: $\rho_k = \begin{cases} \dfrac{-\theta_{12}}{1+\theta_{12}^2} & k=12 \\ 0 & \text{otherwise} \end{cases}$

Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal MA(1)

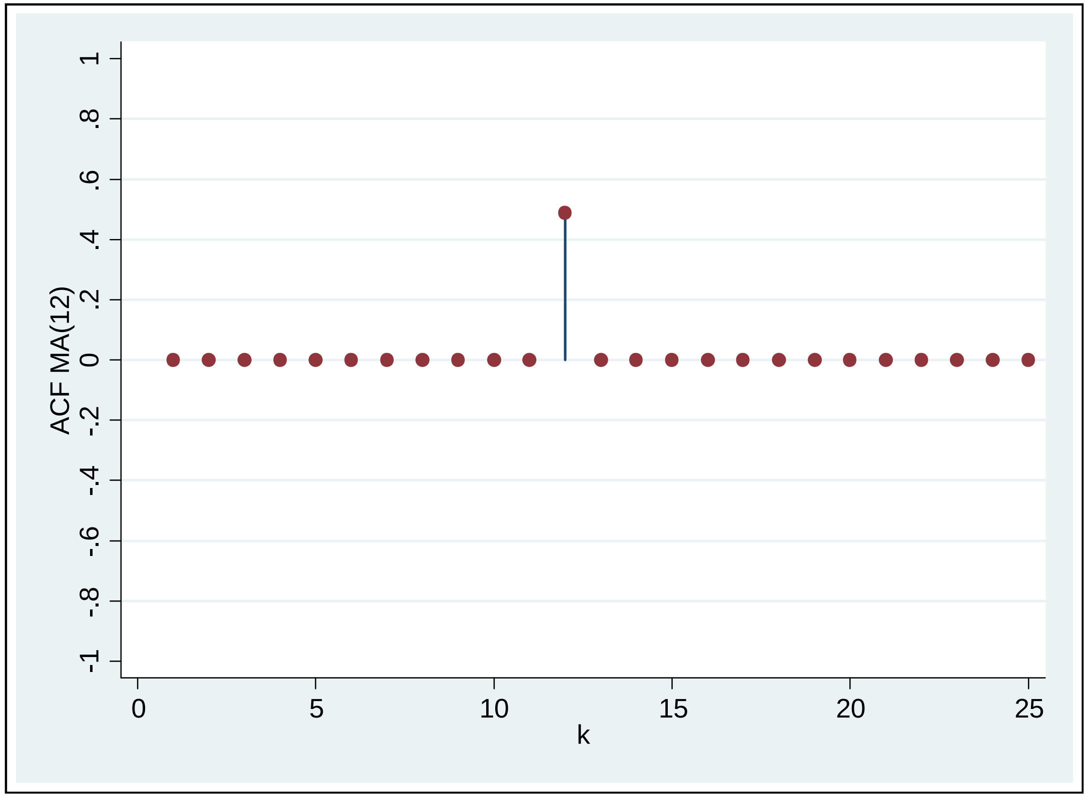
Simulated MA(12) with s = 12 and θ_{12} = -0.8



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal MA(1)

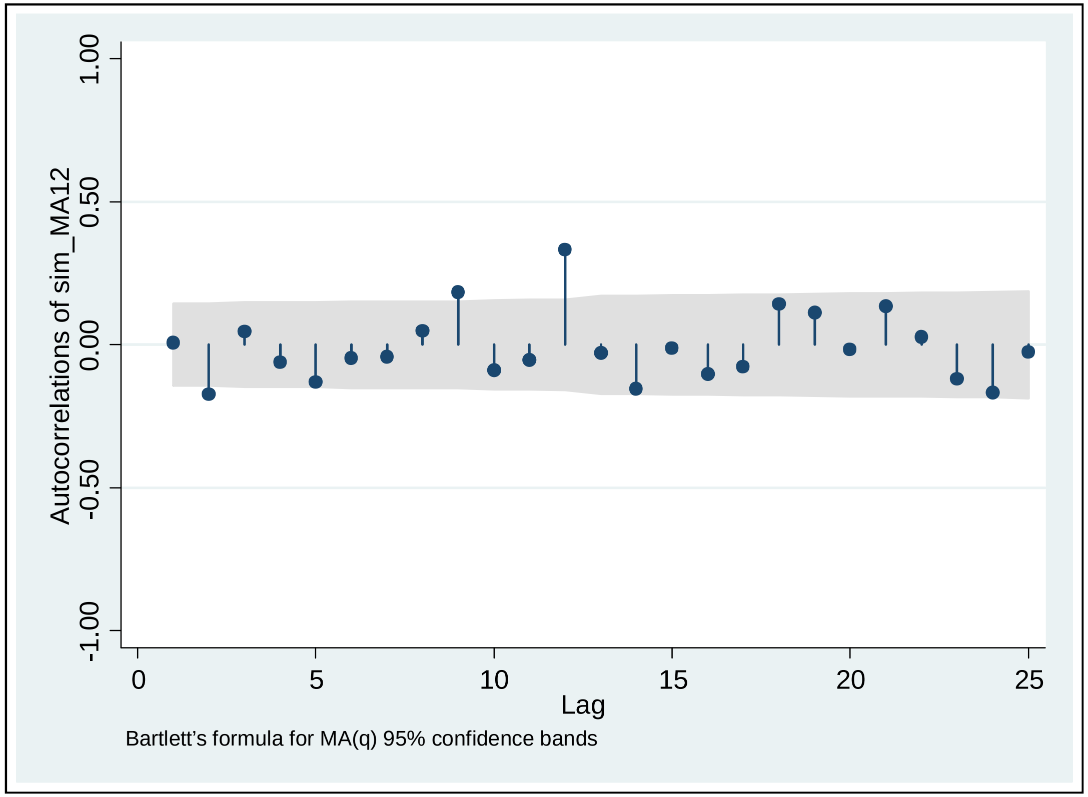
Theoretical ACF of a MA(12) with s = 12 and θ_{12} = -0.8



Harvey (1981) “Time Series Models”, p. 171-185

Purely Seasonal MA(1)

ACF of a simulated MA(12) with s = 12 and θ_{12} = -0.8



Harvey (1981) "Time Series Models", p. 171-185

ACF of Purely Seasonal MA(1)

$$y_t = \left(1 - \theta_1^s L^s\right)\zeta_t$$

Recall: Autocorrelation function for an MA(1) process

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

$$\text{for } s = 4\text{:} \quad \rho_k = \begin{cases} \frac{-\theta_{12}}{1+\theta_{12}^2} & k = 4 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } s = 12\text{:} \quad \rho_k = \begin{cases} \frac{-\theta_{12}}{1+\theta_{12}^2} & k = 12 \\ 0 & \text{otherwise} \end{cases}$$

Harvey (1981) "Time Series Models", p. 171-185

Seasonal ARMA Processes

General Formulation to allow for both AR and MA terms at specified seasonal lags

$$\left(1-\phi_1^s L^s - \ldots - \phi_P^s L^{P_s}\right) y_t = \left(1-\theta_1^s L^s - \ldots - \theta_Q^s L^{Q_s}\right)\zeta_t$$

$$\phi^s\left(L^s\right) y_t = \theta^s\left(L^s\right)\zeta_t$$

with *s* denoting the number of seasons in the year and with ζ_t denoting a white noise disturbance term

→ **pure seasonal ARMA** process of order $(P,Q)_s$

→ **ACF will contain 'gaps'** at non-seasonal lags

Harvey (1981) "Time Series Models", p. 171-185

Primary distinguishing characteristics of theoretical ACF's and PACF's for purely seasonal stationary processes

Process	ACF	PACF
S-AR	Tails off toward zero at lags $k \times s$, $k = 1, 2, \ldots$	Cuts off after lag P_S
S-MA	Cuts off after lag Q_S	Tails off at lags $k \times s$, $k = 1, 2, \ldots$
S-ARMA	Tails off at lags $k \times s$	Tails off at lags $k \times s$

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 158

Seasonal ARMA Processes

A **pure seasonal ARMA** is not appropriate unless seasonal movements are the only predictable feature of the series

→ **Multiplicative** seasonal ARMA

Harvey (1981) "Time Series Models", p. 171-185

Seasonal ARMA Processes

Multiplicative Seasonal ARMA

Replace the white noise disturbance ζ_t term by a non-seasonal ARMA(p,q) process, u_t:

$$\phi(L)u_t = \theta(L)\varepsilon_t$$

Plugged in $\phi^s\left(L^s\right)y_t = \theta^s\left(L^s\right)\zeta_t$ yields

$$\phi^s\left(L^s\right)\phi(L)y_t = \theta^s\left(L^s\right)\theta(L)\varepsilon_t$$

$\rightarrow$ ARMA process of order $(p,q)\times(P,Q)_s$

Harvey (1981) “Time Series Models”, p. 171-185

$$\Phi^s\left(L^s\right)\phi\left(L\right)y_t = \theta^s\left(L^s\right)\theta\left(L\right)\varepsilon_t \quad \text{ARMA}\left(p,q\right)\times\left(P,Q\right)_s$$

Example 1: ARMA(1,0)×(1,0)$_S$

$$\left(1-\phi_1 L\right)\left(1-\Phi_1^4 L^4\right)y_t = \varepsilon_t$$

$$\left(1-\phi_1 L-\Phi_1^4 L^4+\phi_1\Phi_1^4 L^5\right)y_t = \varepsilon_t$$

$$y_t = \phi_1 y_{t-1}+\Phi_1^4 y_{t-4}+\phi_1\Phi_1^4 y_{t-5}+\varepsilon_t$$

May be viewed as

$$y_t = \phi_1 y_{t-1}+\phi_4 y_{t-4}+\phi_5 y_{t-5}+\varepsilon_t$$

with $\phi_5 = -\phi_1\phi_4$

Harvey (1981) "Time Series Models", p. 171-185

$$\cdot\cdot^{s}\left(L^{s}\right)\cdot\cdot\left(L\right)y_{t}=\theta^{s}\left(L^{s}\right)\theta\left(L\right)\varepsilon_{t}\qquad \mathrm{ARMA}\left(p,q\right)\times\left(P,Q\right)_{s}$$

Example 2: ARMA(0,1)×(0,1)$_S$

$$y_t=(1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$y_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_1^s\varepsilon_{t-12}+\theta_1\theta_1^s\varepsilon_{t-13}$$

“(Positive) autocorrelations at the seasonal lags being flanked by (negative) autocorrelations at the ‘satellites’”

Mills, p. 167; Abraham & Ledolter (1983), p. 288 et sqq.

Example 2: ARMA(0,1)$\times$(0,1)$_S$

$$y_t = (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1^s\varepsilon_{t-12} + \theta_1\theta_1^s\varepsilon_{t-13}$$

It can be sown that

$$Var(y_t) = \sigma_\varepsilon^2\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right)$$

$$Cov(y_t, y_{t-1}) = -\theta_1\left(1-\theta_1^{s\,2}\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s+1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s}) = -\theta_1^s\left(1-\theta_1^2\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s-1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = 0 \qquad \text{for} \qquad k \neq 0, 1, s-1, s, s+1$$

Tsay, p. 75

Example 2: ARMA(0,1)×(0,1)$_S$

$$y_t = (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1^s\varepsilon_{t-12} + \theta_1\theta_1^s\varepsilon_{t-13}$$

$$Var(y_t) = \sigma_\varepsilon^2\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right)$$

$$Cov(y_t, y_{t-1}) = -\theta_1\left(1-\theta_1^{s\,2}\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s+1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s}) = -\theta_1^s\left(1-\theta_1^2\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s-1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = 0$$

for $\quad k \neq 0, 1, s-1, s, s+1$

$$\rho_1 = -\theta_1/\left(1+\theta_1^2\right)$$

$$\rho_{s-1} = \theta_1\theta_1^s/\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right) = \rho_1\cdot\rho_s$$

$$\rho_s = -\theta_1^s/\left(1+\theta_1^{s\,2}\right)$$

$$\rho_{s-1} = \theta_1\theta_1^s/\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right) = \rho_1\cdot\rho_s$$

$$\rho_k = 0 \qquad \text{otherwise}$$

Tsay, p. 75

Example 2: ARMA(0,1)×(0,1)$_S$

$$y_t = (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1^s\varepsilon_{t-12} + \theta_1\theta_1^s\varepsilon_{t-13}$$

$$Var(y_t) = \sigma_\varepsilon^2\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right)$$

$$Cov(y_t, y_{t-1}) = -\theta_1\left(1+\theta_1^{s\,2}\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s+1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s}) = -\theta_1^s\left(1+\theta_1^2\right)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-s-1}) = \theta_1\theta_1^s\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = 0$$

for $\quad k \neq 0,1,s-1,s,s+1$

Like MA(1)

„Interaction“ at the „satellites“

$$\boxed{\rho_1 = -\theta_1/\left(1+\theta_1^2\right)}$$

$$\rho_{s-1} = \theta_1\theta_1^s/\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right) = \boxed{\rho_1\cdot\rho_s}$$

$$\boxed{\rho_s = -\theta_1^s/\left(1+\theta_1^{s\,2}\right)}$$

Like MA(1)$_S$

$$\rho_{s-1} = \theta_1\theta_1^s/\left(1+\theta_1^{s\,2}\right)\left(1+\theta_1^2\right) = \boxed{\rho_1\cdot\rho_s}$$

$$\rho_k = 0 \qquad \text{otherwise}$$

Tsay, p. 75

$$\cdot\cdot^{s}\left(L^{s}\right)\cdot\cdot\left(L\right)y_t = \theta^{s}\left(L^{s}\right)\theta\left(L\right)\varepsilon_t \quad \text{ARMA}\left(p,q\right)\times\left(P,Q\right)_s$$

Example 2: ARMA(0,1)$\times$(0,1)$_S$

$$y_t = (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1^s\varepsilon_{t-12} + \theta_1\theta_1^s\varepsilon_{t-13}$$

Note: this is the stationary part of the famous “airline model” **ARIMA(0,1,1)$\times$(0,1,1)$_S$** (see example below)

$$(1-L^{12})(1-L)y_t = (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

$$\phi^{s}\left(L^{s}\right)\phi\left(L\right)y_{t} = \theta^{s}\left(L^{s}\right)\theta\left(L\right)\varepsilon_{t} \quad \text{ARMA}\left(p,q\right)\times\left(P,Q\right)_{s}$$

Example 3: ARMA(0,1)×(1,1)$_S$

$$\left(1-\phi_{1}^{s}L^{12}\right)y_{t} = (1-\theta_{1}^{s}L^{12})(1-\theta_{1}L)\varepsilon_{t}$$

$$y_{t} = \phi_{1}^{s}y_{t-12} + \varepsilon_{t} - \theta_{1}\varepsilon_{t-1} - \theta_{1}^{s}\varepsilon_{t-12} + \theta_{1}\theta_{1}^{s}\varepsilon_{t-13}$$

A General Class of Models

Box and Jenkins propose that the conventional and seasonal differencing operators be applied until the series is stationary and that this stationary series be modeled by a multiplicative seasonal ARMA

$$\phi^s\left(L^s\right)\phi\left(L\right)\Delta^d \Delta_s^D y_t = \theta^s\left(L^s\right)\theta\left(L\right)\varepsilon_t$$

where D and d are integers denoting the number of times the seasonal and first difference operators are applied respectively.

Multiplicative Seasonal ARIMA of order $\left(p,d,q\right)\times\left(P,D,Q\right)_s$

Harvey (1981) “Time Series Models”, p. 171-185

Multiplicative Seasonal ARIMA of order $(p,d,q)\times(P,D,Q)_s$

$$\cdot\cdot^{\,s}\left(L^{s}\right)\cdot\cdot\left(L\right)\Delta^{d}\Delta_{s}^{D}y_{t}=\theta^{s}\left(L^{s}\right)\theta\left(L\right)\varepsilon_{t}$$

1. **Determine *d* and *D***
   high sample ACFs that slowly die out (at multiples of *s*) require (seasonal) differencing. ACF of differenced series should look stationary $(1-L)^{d}(1-L^{s})^{D}y_{t}$

2. **Determine *p*,*q* and *P*,*Q***
   by studying ACF, PACF of $(1-L)^{d}(1-L^{s})^{D}y_{t}$

Harvey (1981) “Time Series Models”, p. 171-185

Multiplicative Seasonal ARIMA Processes

Multiplicative Seasonal ARIMA of order $(p,d,q)\times(P,D,Q)_s$

$$\phi^s(L^s)\phi(L)\Delta^d\Delta_s^D y_t = \theta^s(L^s)\theta(L)\varepsilon_t$$

1. Determine *d* and *D*

"It is difficult (for a nonstationary series) to isolate any seasonal pattern as all autocorrelations are dominated by the effect of the nonseasonal unit root.“

Mills, p. 167

This suggests to determine *d* first. However, (ADF) unit root tests require white noise „error term“. Box and Jenkins strategy: Difference until correlogram looks like that from a stationary process.

„Some experimentation with various combinations of first differences and seasonal differences may be necessary."

Harvey (1981) "Time Series Models", p. 171-185

Example:

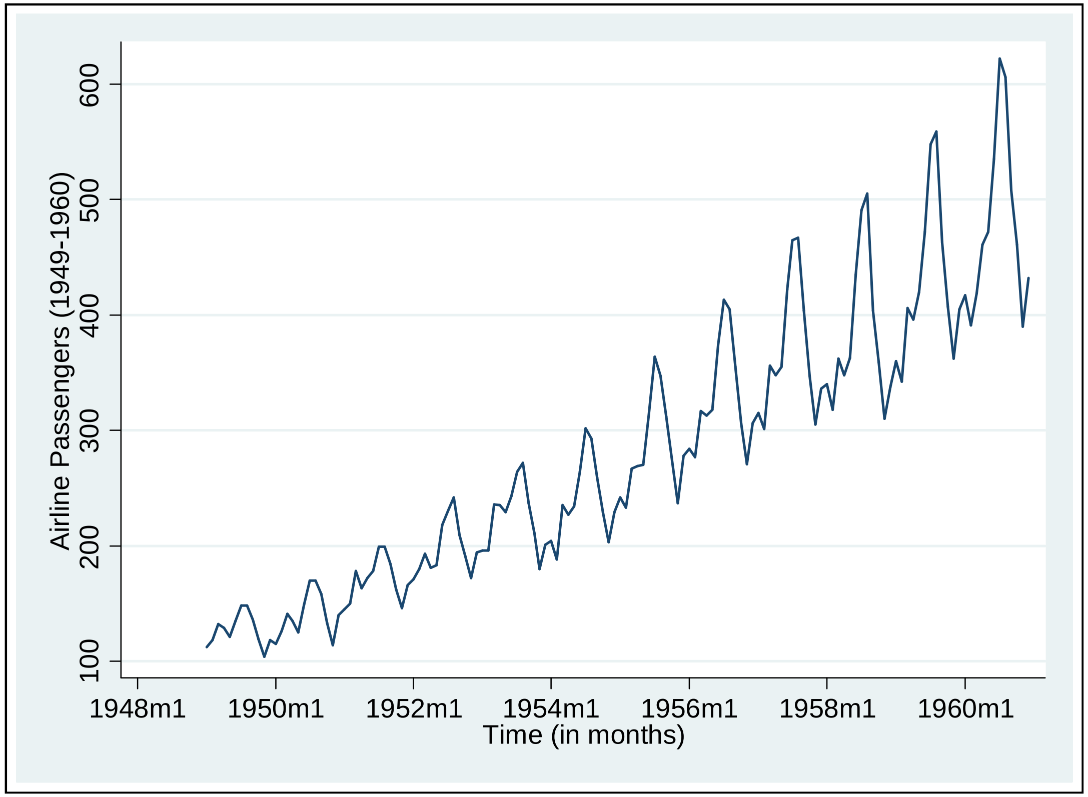
Airline Passengers

Monthly totals of passengers (in thousands) over the period January 1949 to December 1960

Box, Jenkins “Time Series Analysis forecasting and control” p.300-322

Example: Airline Passenger Series

Original series y_t

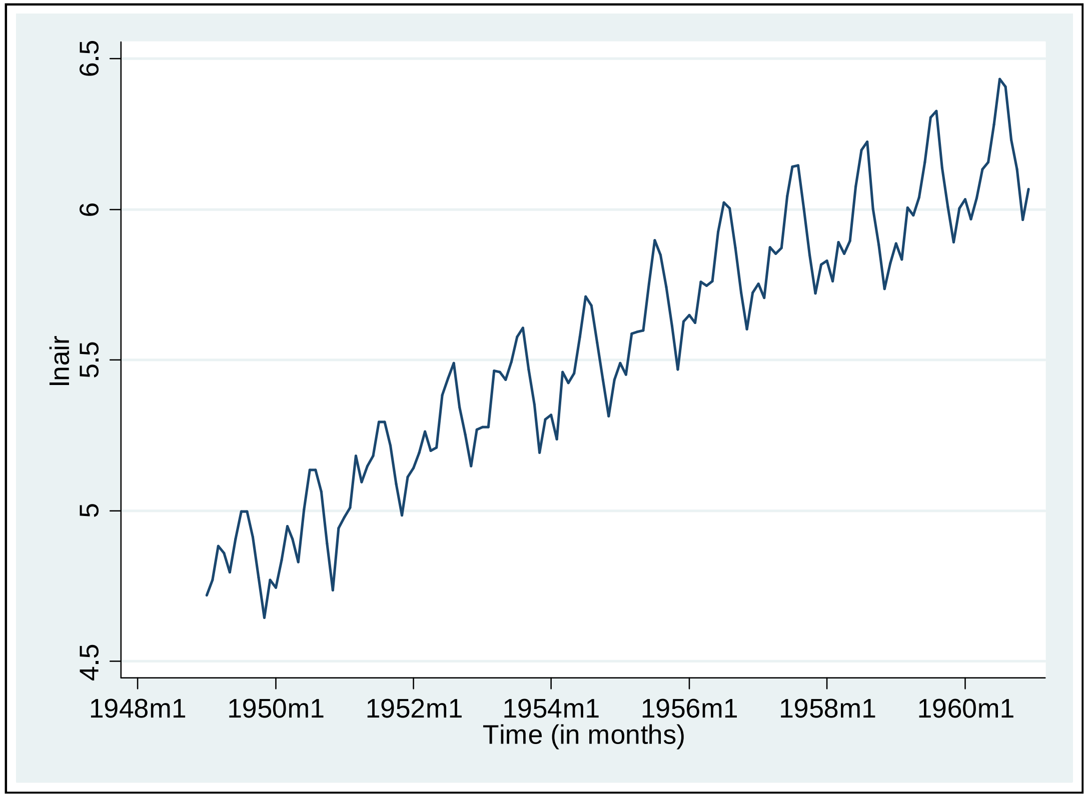


The series shows a marked seasonal pattern since travel is at its highest in the late summer month, while a secondary peak occurs in the spring.

Note the increasing variance.

Example: Airline Passenger Series

Logarithms of the original series $\ln y_t$



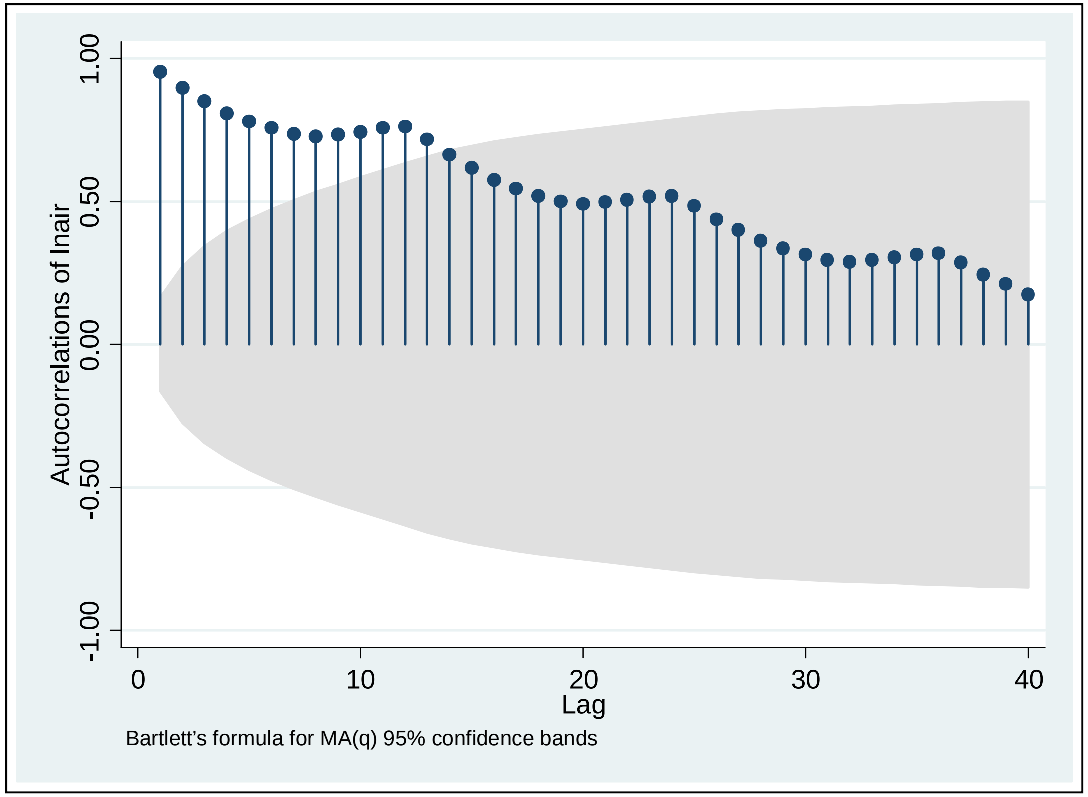
Exponential trend in original series is transformed into linear trend in the log series.

Log- transformation stabilizes the variance ("makes the variance stationary").

Log- transformation must be "undone" for forecasting the series in original units. Careful here! (see below)

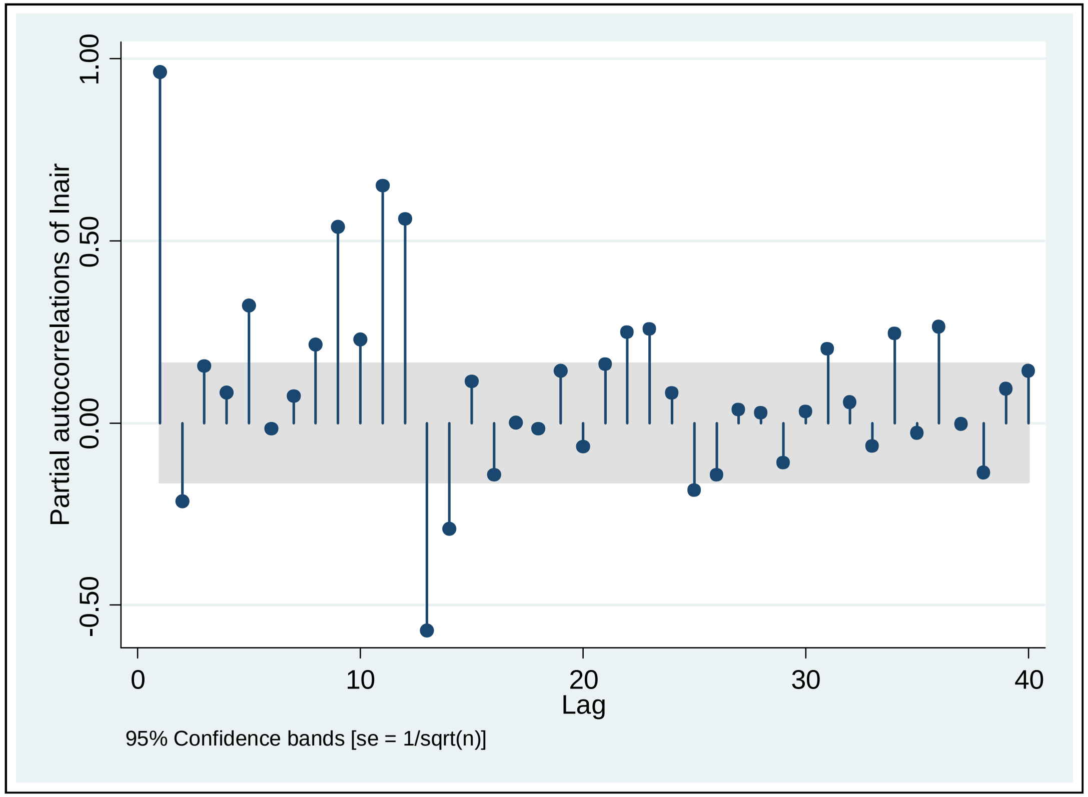
Example: Airline Passenger Series

ACF of logarithms of the original series



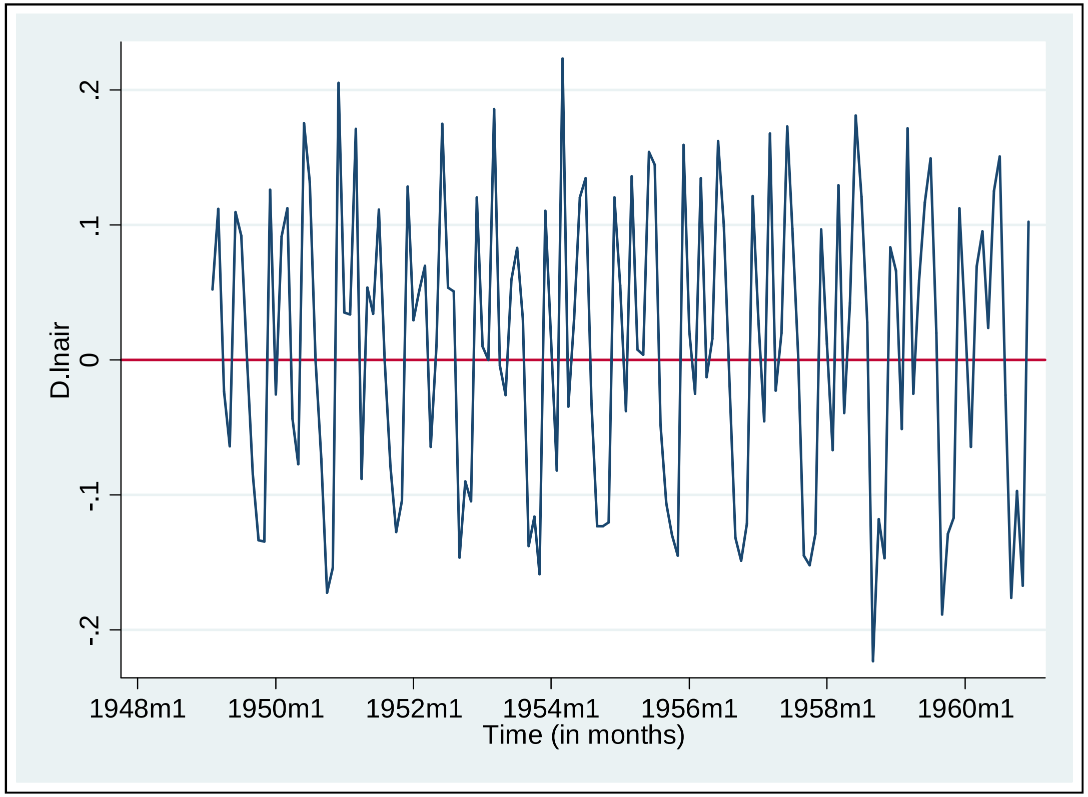
Example: Airline Passenger Series

PACF of logarithms of the original series



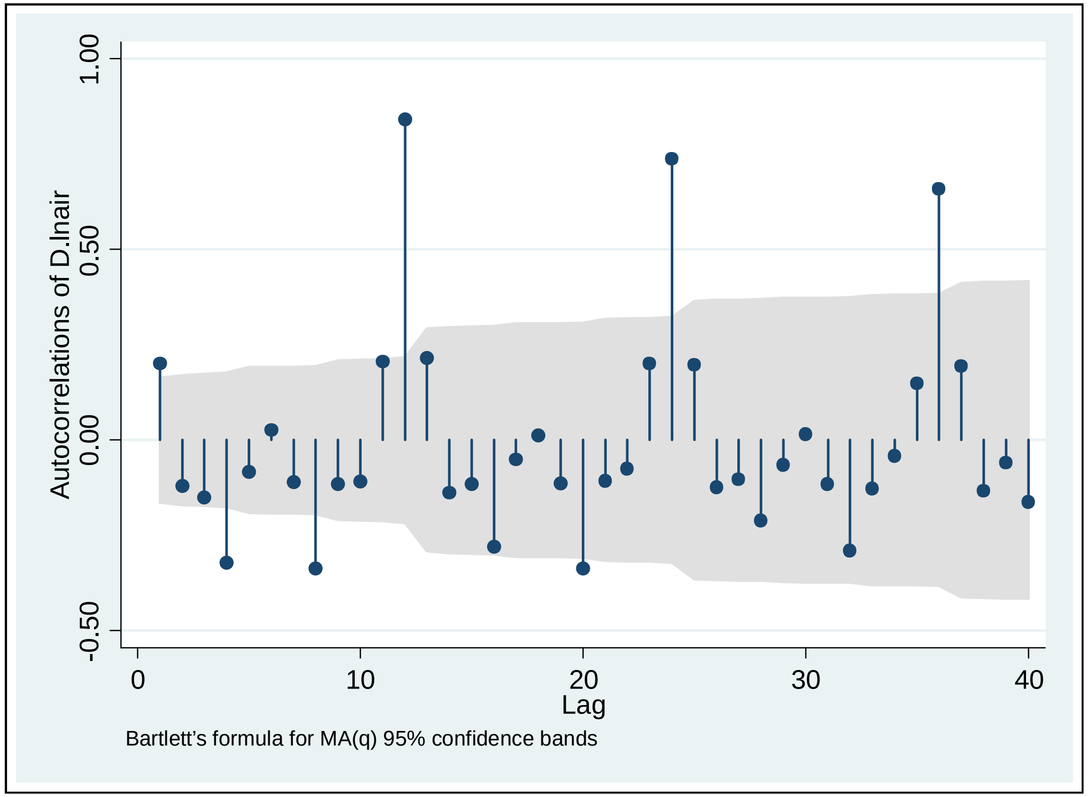
Example: Airline Passenger Series

First (non-seasonal) differences $(1-L)\ln y_t = \Delta \ln y_t = \ln y_t - \ln y_{t-1}$



Example: Airline Passenger Series

ACF of the first (non-seasonal) differences

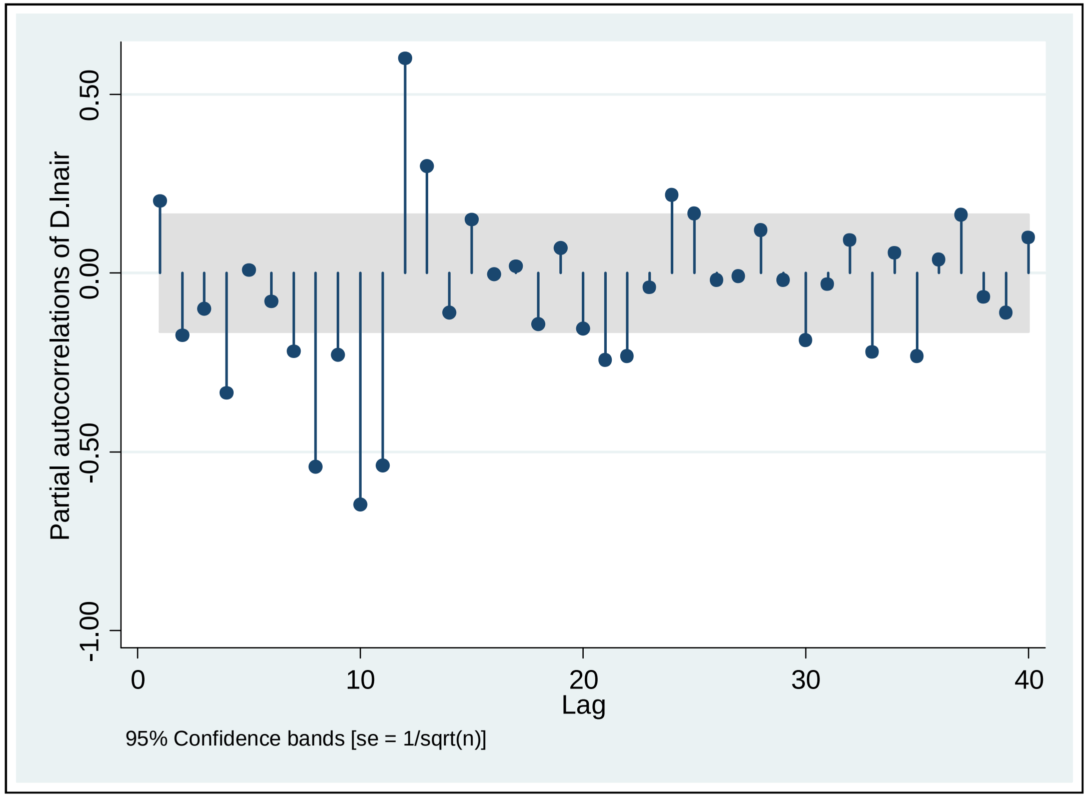


Slowly declining ACF at seasonal lags

=> permanent, non-stationary seasonality that should be removed by seasonal differencing.

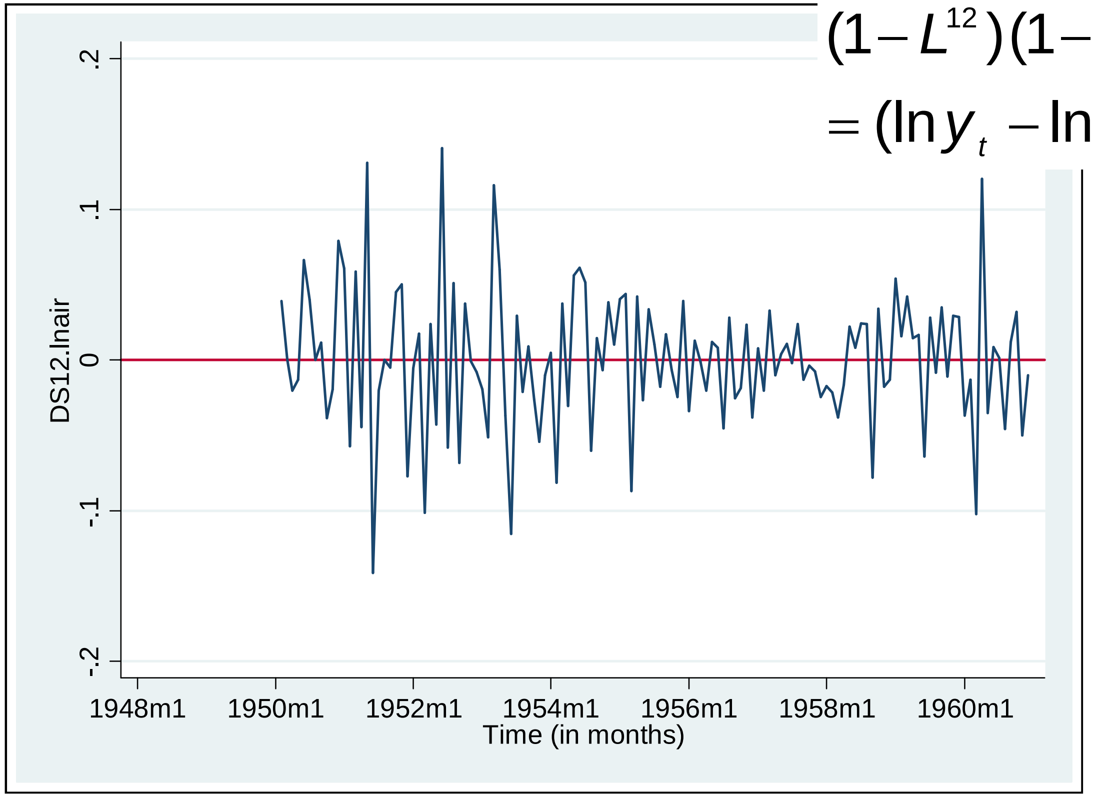
Example: Airline Passenger Series

PACF of the first (non-seasonal) differences



Example: Airline Passenger Series

Seasonal difference of the differenced series

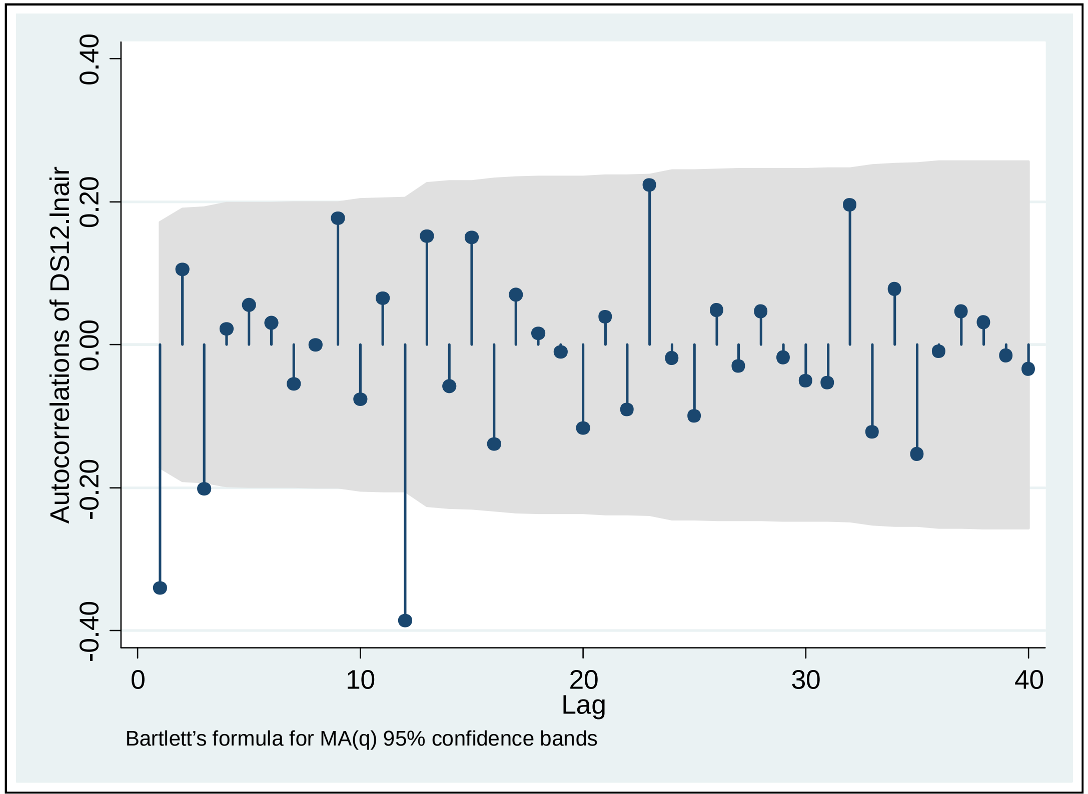


$$(1-L^{12})(1-L)y_t = \Delta\Delta_{12}y_t$$
$$= (\ln y_t - \ln y_{t-1}) - (\ln y_{t-12} - \ln y_{t-13})$$

Only 131 of 144 observations are left

Example: Airline Passenger Series

ACF of seasonal difference of the differenced series



Two spikes stand out:
at lags 1 and 12

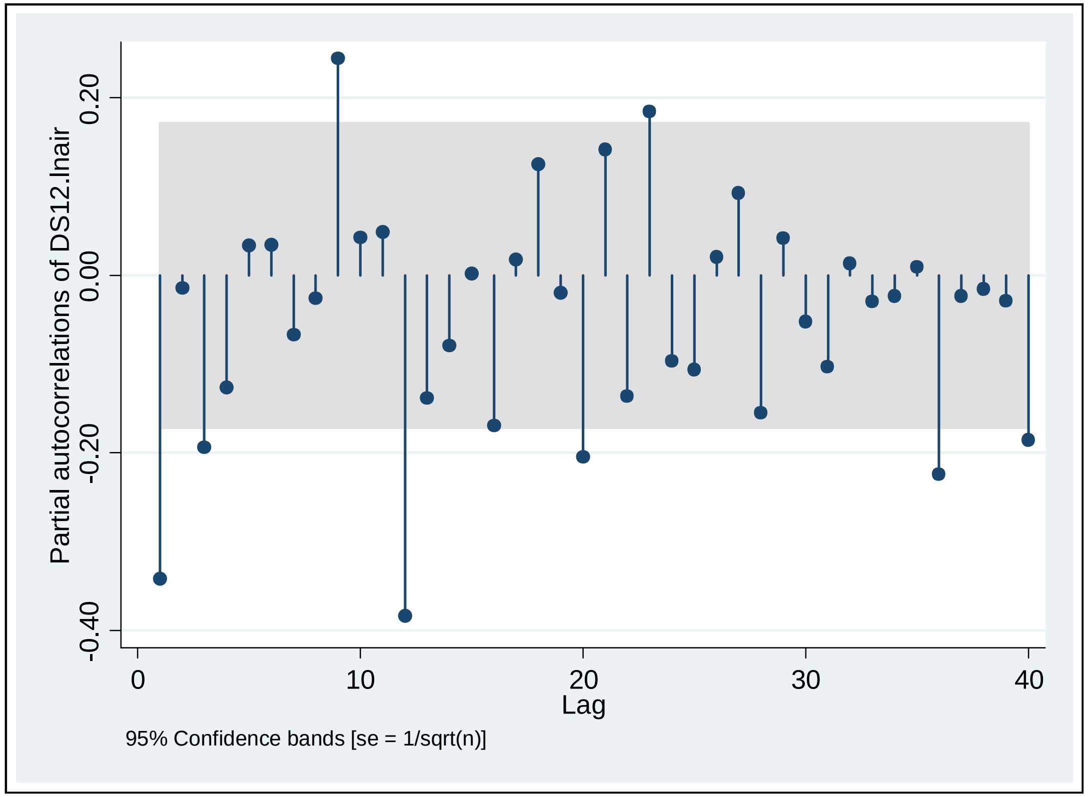
$\Rightarrow$non-seasonal MA(1)
and seasonal MA(1)

$\Rightarrow$ May try

$$(1-L^{12})(1-L)\ln y_t$$
$$= (1-\theta_1^s L^{12})(1-\theta_1 L)\varepsilon_t$$

Example: Airline Passenger Series

PACF of seasonal difference of the differenced series



Example: Airline Passengers

Identification

- substantial peak in the ACF at lag $k = 1$, indicating a possible moving average of order $q = 1$
- substantial peak in the ACF at lag $k = 12$, indicating a possible moving average of order $Q = 12$

→ **Candidate model:**

- ARIMA (0,1,1) x $(0,1,1)_{12}$ with y_t in logarithms.

$$\Delta\Delta_{12} \ln y_t = (1-\theta_1 L)(1-\theta_{12} L^{12})\varepsilon_t$$

Example: Airline Passenger Series

Estimation of ARIMA (0,1,1) x $(0,1,1)_{12}$

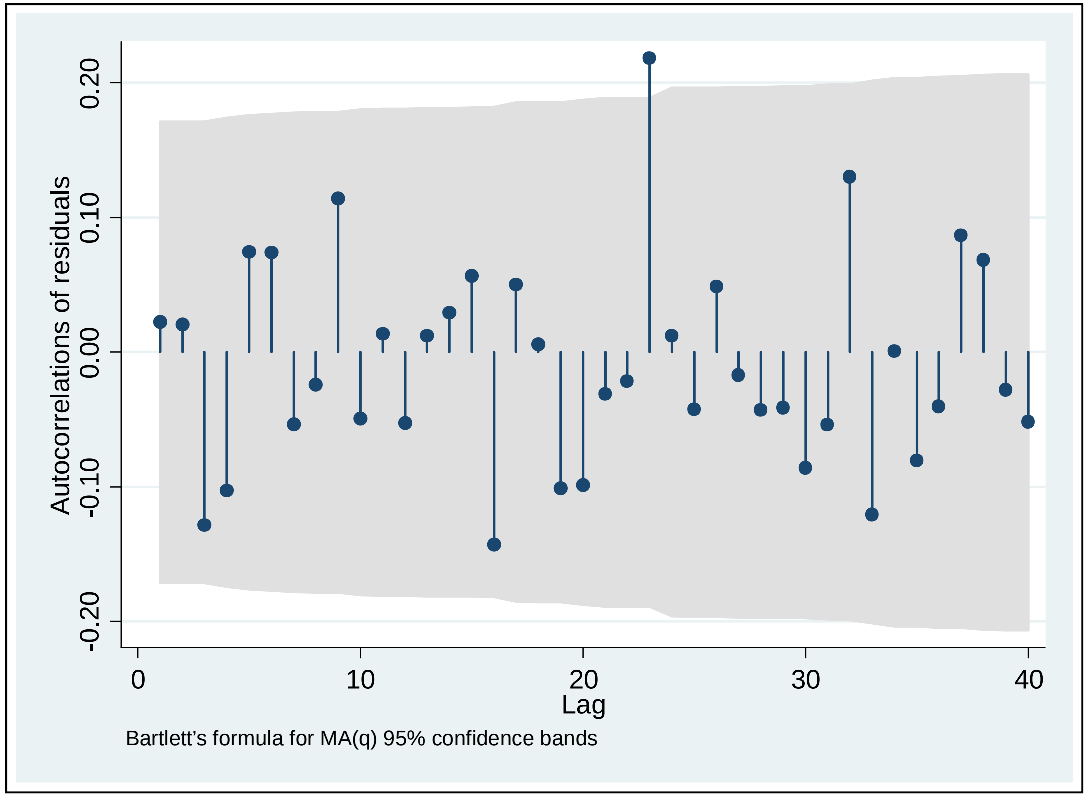
```
. arima lnair, arima(0,1,1) sarima(0,1,1,12) noconstant
[...]
ARIMA regression
Sample:  1950m2 to 1960m12                        Number of obs      =       131
                                                  Wald chi2(2)       =     84.53
Log likelihood =  244.6965                        Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
  DS12.lnair |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -.4018324    .0730307    -5.50   0.000    -.5449698   -.2586949
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.5569342    .0963129    -5.78   0.000     -.745704   -.3681644
-------------+----------------------------------------------------------------
      /sigma |   .0367167    .0020132    18.24   0.000     .0327708    .0406625
------------------------------------------------------------------------------
```

$$(1-L^{12})(1-L)\ln y_t = (1-0.55L^{12})(1-0.4L)\varepsilon_t$$

$$\ln y_t = \ln y_{t-1} + \ln y_{t-12} - \ln y_{t-13} + \varepsilon_t - 0.4\varepsilon_{t-1} - 0.55\varepsilon_{t-12} + 0.22\varepsilon_{t-13}$$

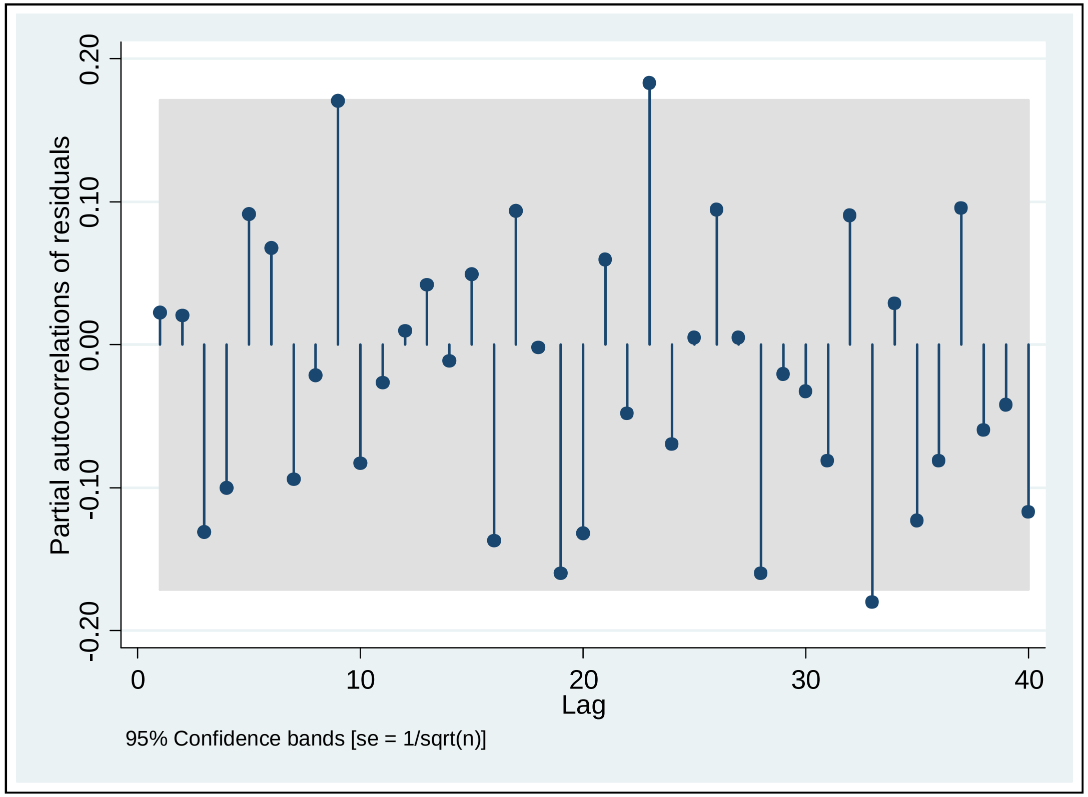
Example: Airline Passenger Series

ACF of residuals from the ARIMA (0,1,1) x $(0,1,1)_{12}$ model



Example: Airline Passenger Series

PACF of residuals from the ARIMA (0,1,1) x $(0,1,1)_{12}$ model



Example: Airline Passenger Series

Diagnostic checking of ARIMA (0,1,1) x $(0,1,1)_{12}$

Joint Hypothesis Test

H_0: All autocorrelation coefficients are zero

Box and Ljung (refined test)

$$Q = T(T+2)\sum_{k=1}^{K} \frac{1}{T-K}\hat{\rho}_k^2 \quad \sim \chi^2$$

with $K - p - q$ degrees of freedom

Example: Airline Passenger Series

Diagnostic checking of ARIMA (1,1,0) x $(0,1,1)_{12}$

```
. corrgram residuals, lags(24)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
[...]
12       -0.0526   0.0096   8.4712  0.7473          |                  |
[...]
24        0.0121  -0.0696    23.62  0.4835          |                  |

. di 1-chi2(10, 8.4712)
.58291173
. di 1-chi2(22, 23.62)
.36745865
```

Example: Airline Passenger Series

Forecasting with ARIMA (0,1,1) x $(0,1,1)_{12}$

Forecasts at the arbitrarily selected origin, Dec. 1959

Reestimating the model using data from 1950m2 to 1959m12:

```
. arima lnair if time <=-1, arima(0,1,1) sarima(0,1,1,12) noconstant
ARIMA regression
Sample:  1950m2 to 1959m12                      Number of obs      =       119
                                                Wald chi2(2)       =     68.54
Log likelihood =  223.6266                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
  DS12.lnair |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -.3484396   .0810301    -4.30   0.000    -.5072558   -.1896235
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.5622757   .0944329    -5.95   0.000    -.7473608   -.3771906
-------------+----------------------------------------------------------------
      /sigma |   .0362307   .0021329    16.99   0.000     .0320502    .0404111
----------------------------------------------------------------------------
```

Example: Airline Passenger Series

Forecasting with ARIMA (1,1,0) x (0,1,1)$_{12}$

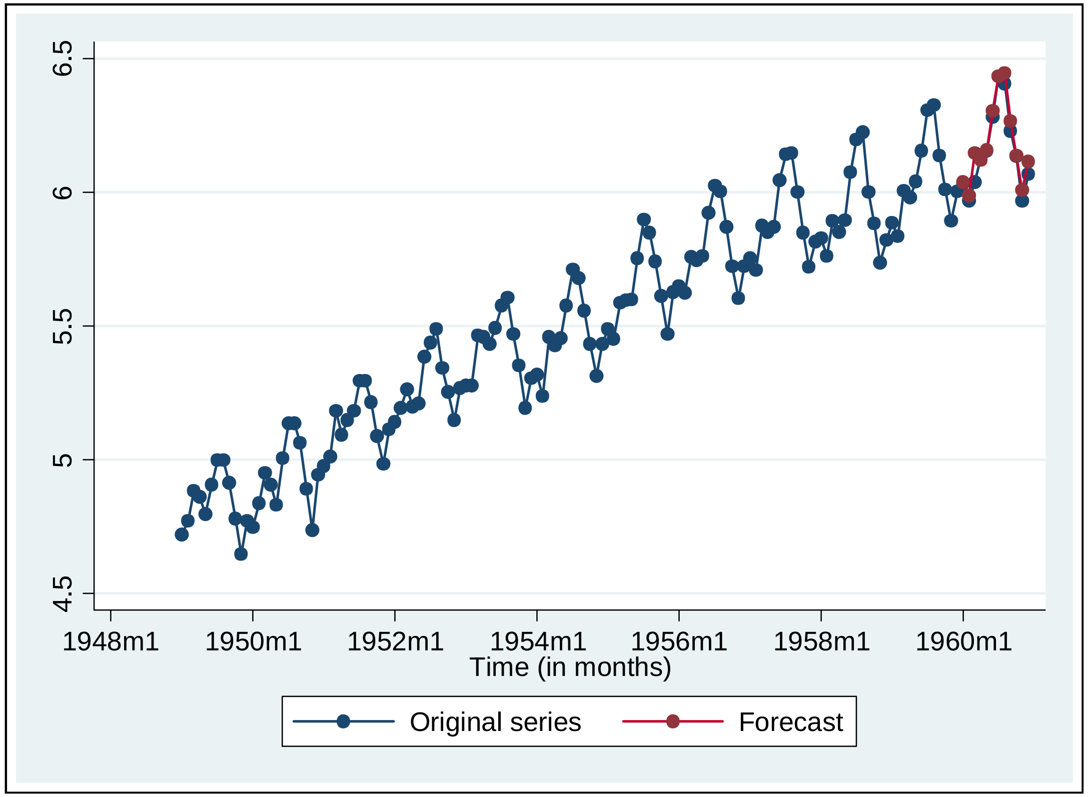
Forecasts at the arbitrarily selected origin, Dec. 1959:

$$\ln y_{T+l|} = \ln y_{T+l-1|} + \ln y_{T+l-12} - \ln y_{T+l-13}$$
$$+ \varepsilon_{T+l} - 0.35\varepsilon_{T+l-1} - 0.56\varepsilon_{T+l-12} + 0.2\varepsilon_{T+l-13}$$

$$\ln \tilde{y}_{T+1|T} = \ln y_T + \ln y_{T-11} - \ln y_{T-12|T} \quad + \underbrace{\tilde{\varepsilon}_{T+1|T}}_{=0} - 0.35\varepsilon_T - 0.56\varepsilon_{T-11|} + 0.2\varepsilon_{T-12}$$
$$= 6 + 5.89 - 5.82 - 0.35 \cdot 0.0172 - 0.56 \cdot 0.0314 - 0.2 \cdot 0.039$$
$$= 6.04$$

Example: Airline Passenger Series

Forecasting with ARIMA (0,1,1) x (0,1,1)$_{12}$ model



Harvey (1981) “Time Series Models”, p. 171-185

Example: Airline Passenger Series

Forecasting with ARIMA (1,1,0) x (0,1,1)$_{12}$

1-step-ahead forecast of **log** passengers: $\ln \tilde{y}_{T+1|T} = 6.04$

To obtain 1-step-ahead forecast of passengers, note that

in general $e^{E[\ln Y]} \neq E[Y]$

If $\ln Y \sim N(\mu, \sigma^2) \Rightarrow Y \sim \text{lognormal}(\mu, \sigma^2)$

and

$$E[Y] = e^{E[\ln Y]} e^{0.5\sigma^2} = e^{\mu} e^{0.5\sigma^2} = e^{\mu + 0.5\sigma^2}$$

Example: Airline Passenger Series

Forecasting with ARIMA (1,1,0) x (0,1,1)$_{12}$

1-step-ahead forecast of **log** passengers: $\ln \tilde{y}_{T+1|T} = 6.04$

To obtain 1-step-ahead forecast of passengers use estimated version of

$$E[Y_{T+1} \mid \Omega_T] = e^{E[\ln Y_{T+1} | \Omega_T]} e^{0.5\sigma^2}$$

```
-------------------------------------------------------------------------------
-------------+-----------------------------------------------------------------
      /sigma |   .0362307   .0021329    16.99   0.000     .0320502    .0404111
------------------------------------------------------------------------------
```

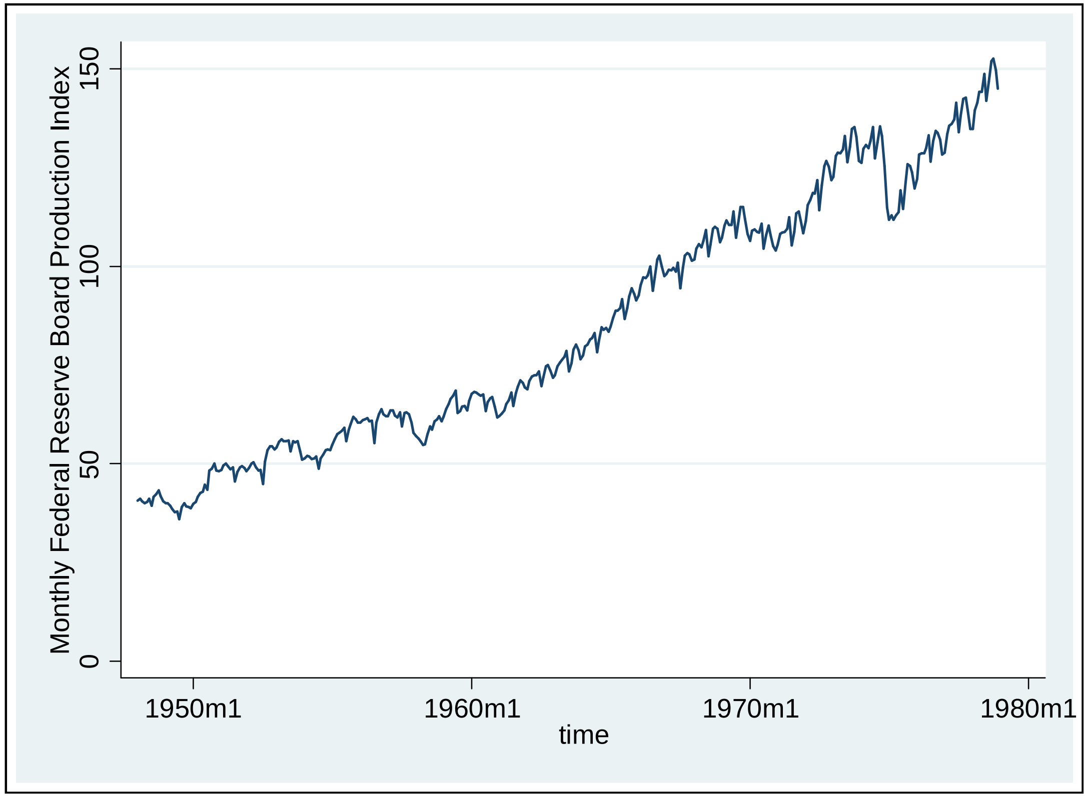
Example:

Federal Reserve Board Production Index

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

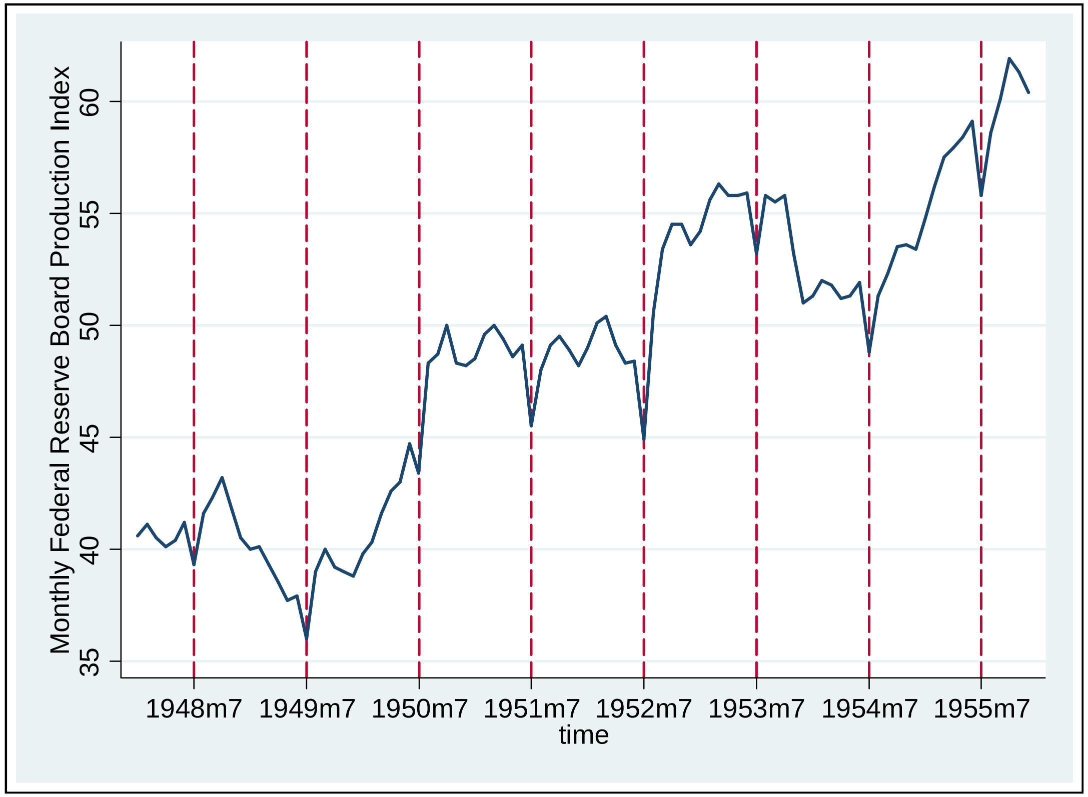
Complete original series



Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

First eight years of the original series

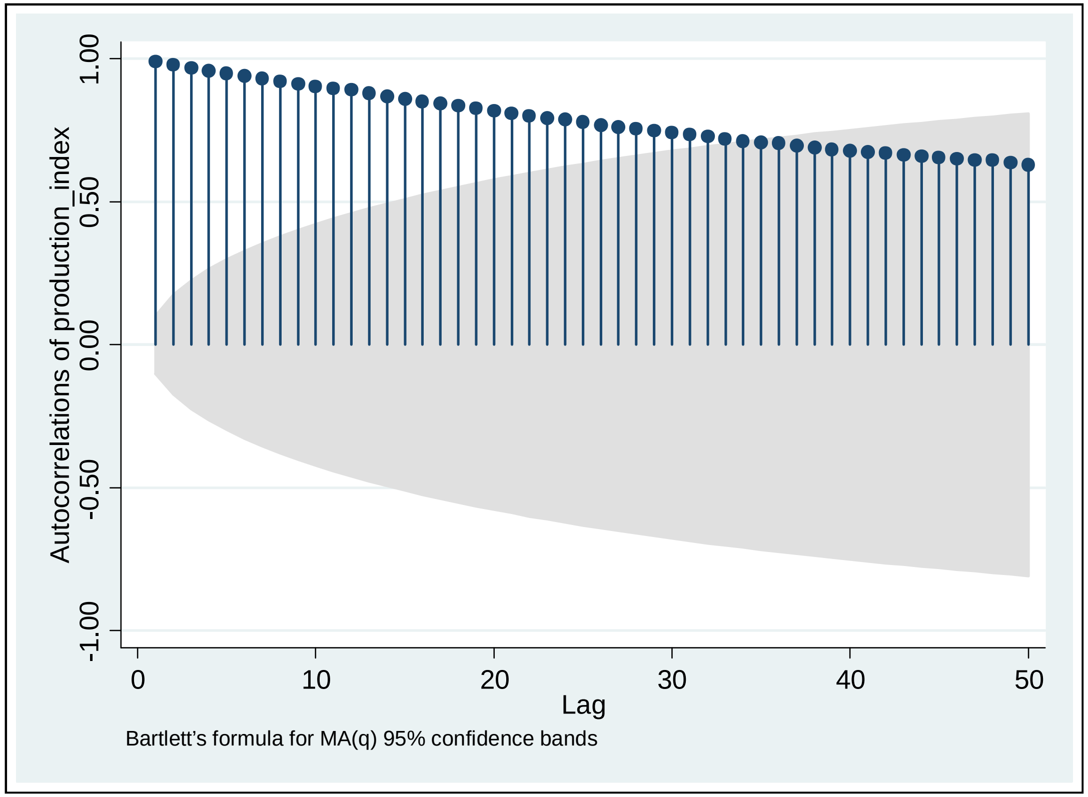


Note for example the troughs in July.

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

ACF of the original series

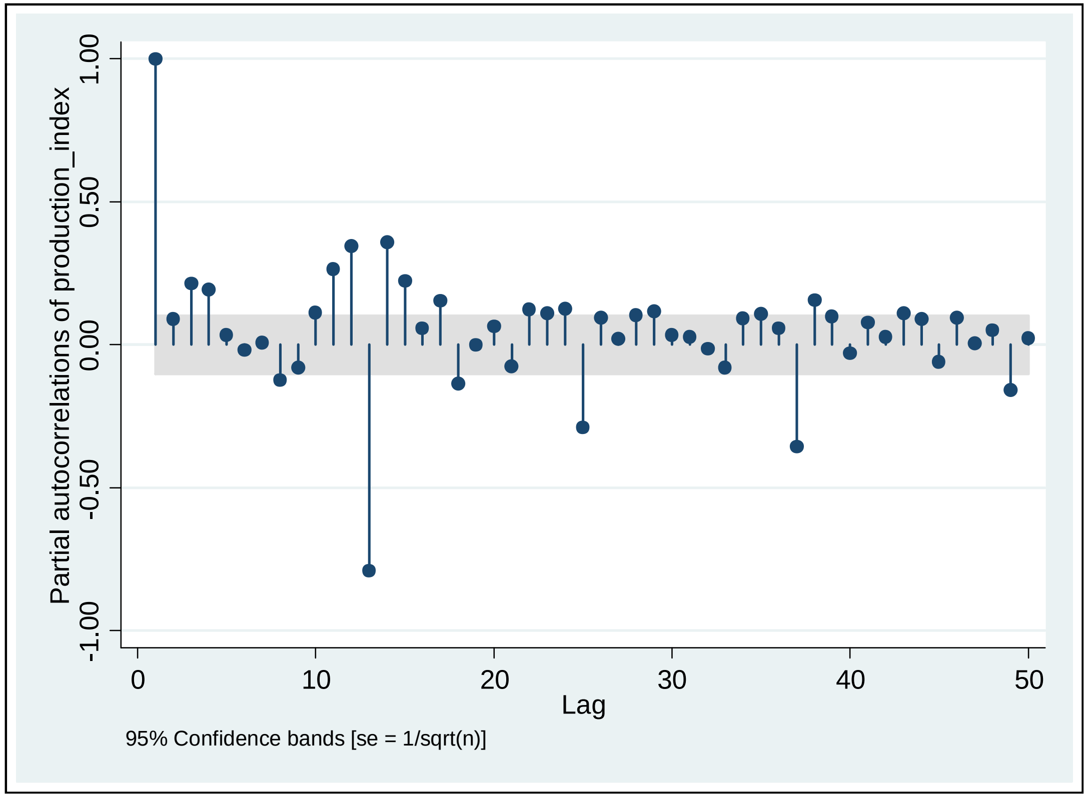


Note the slow decay in the ACF, indicating nonstationary behavior.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

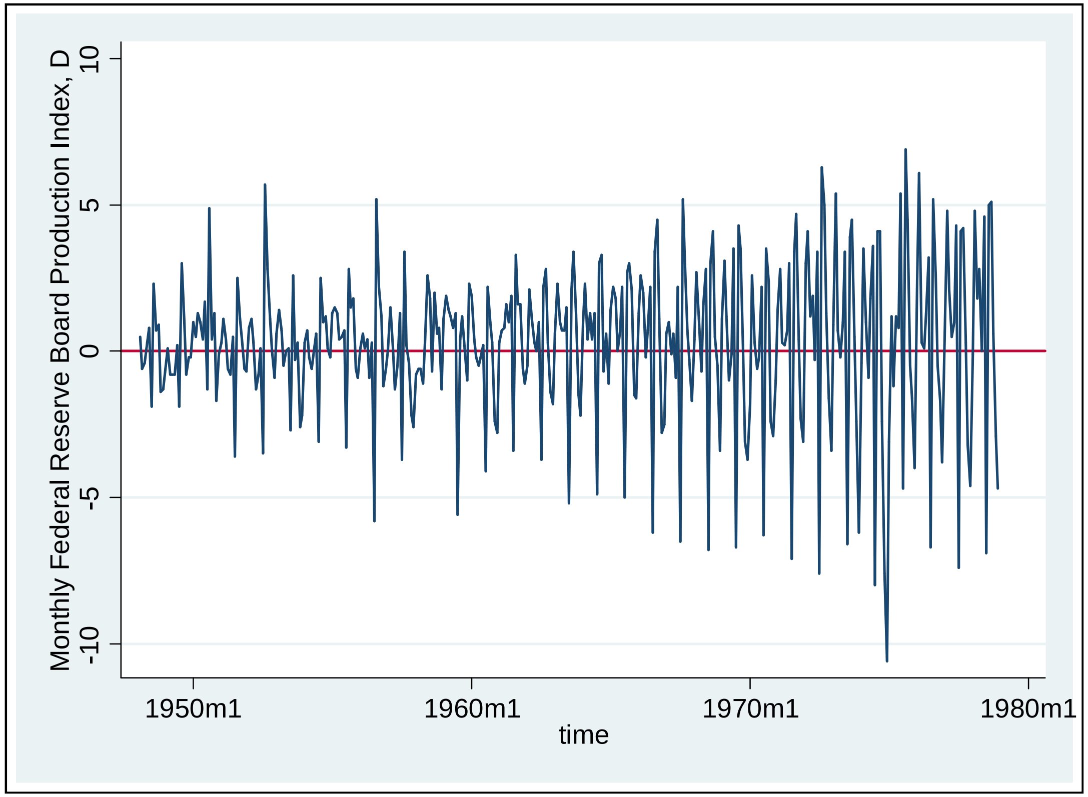
PACF of the original series



Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

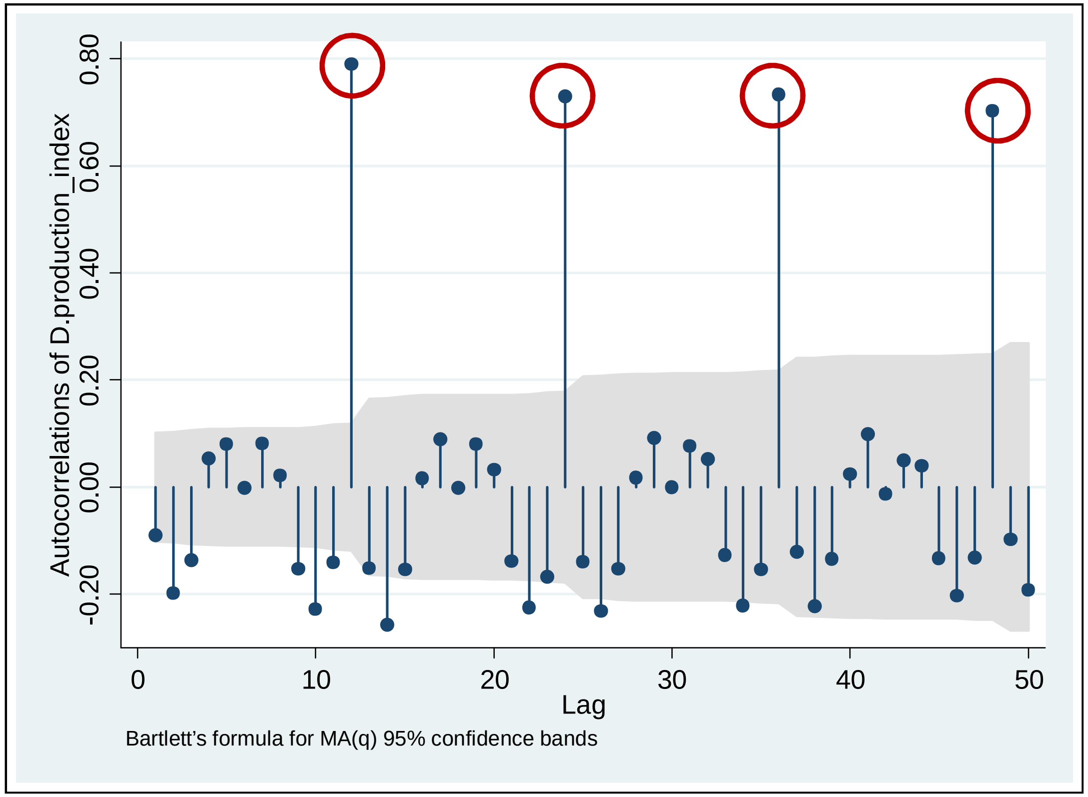
First (non-seasonal) differences $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$



Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

ACF of the first (non-seasonal) differences $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$

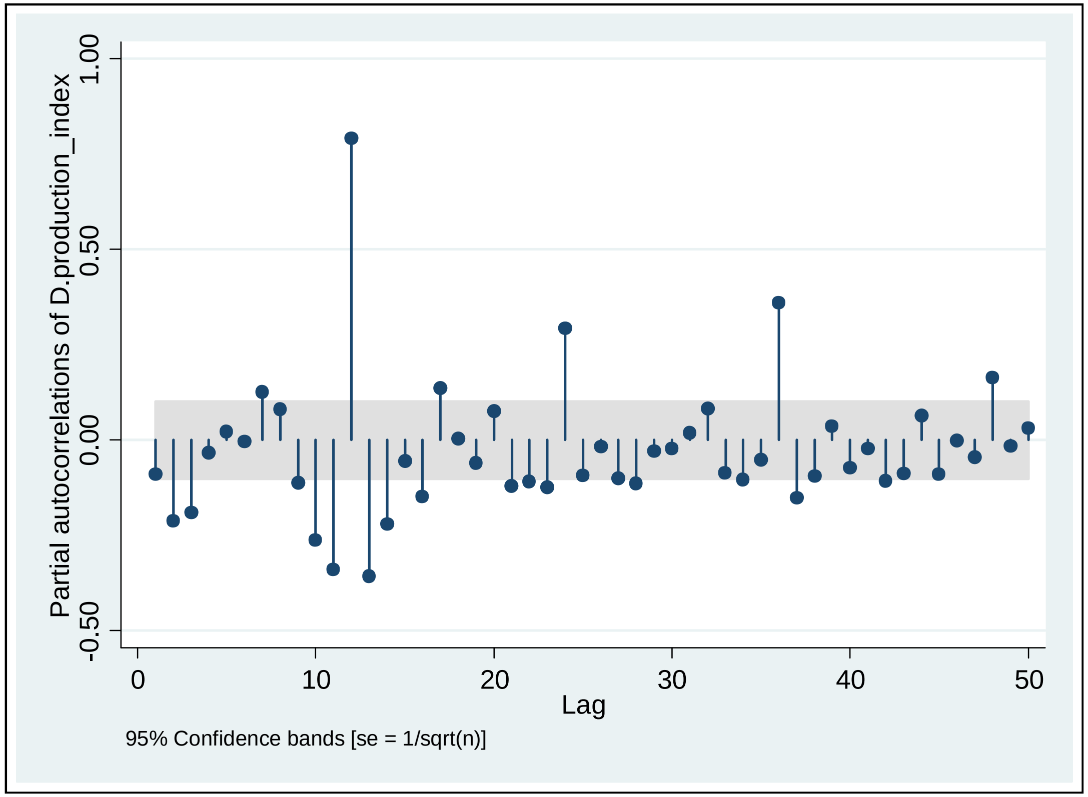


Noting the peaks at 12, 24, 36, and 48 with relatively slow decay suggested a seasonal difference.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

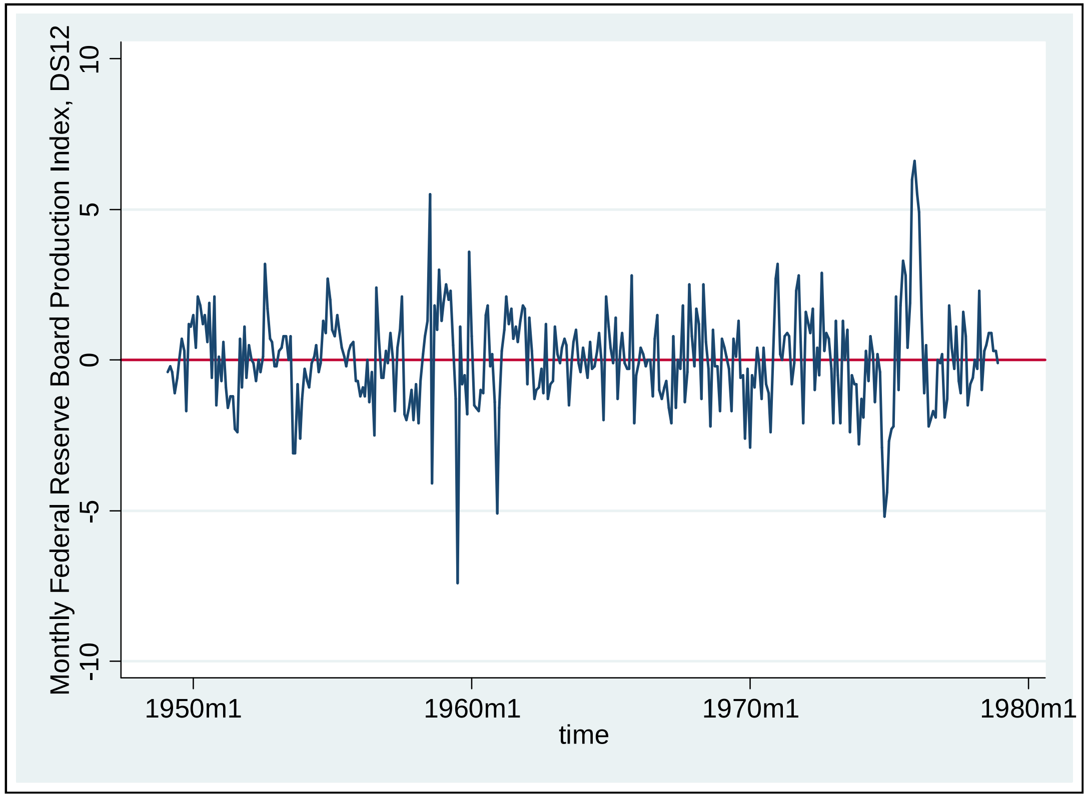
PACF of the first (non-seasonal) differences $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$



Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

Seasonal difference of the differenced series

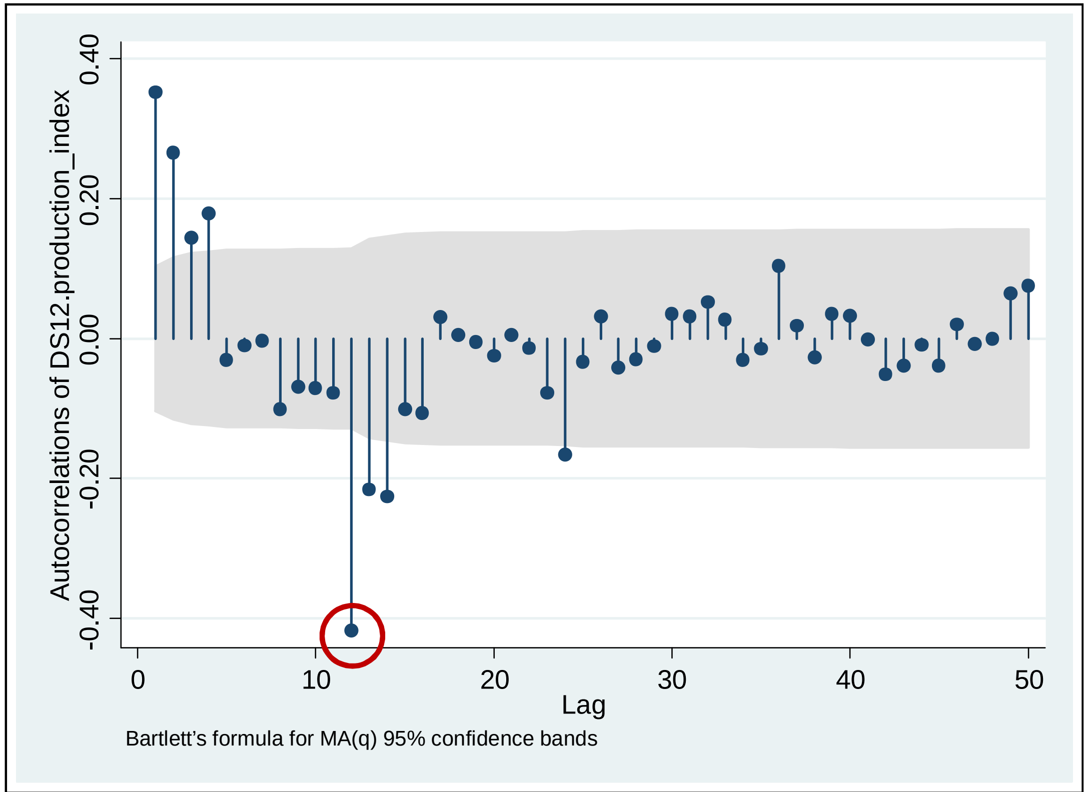


$$\Delta^{d=1}\Delta_{s=12}^{D=1}y_t = (1-L)(1-L^{12})y_t$$
$$= (1-L)(y_t - y_{t-12})$$
$$= (y_t - y_{t-12}) - (y_{t-1} - y_{t-13})$$
$$= (y_t - y_{t-1}) - (y_{t-12} - y_{t-13})$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

ACF of seasonal difference of the differenced series $(1-L)(1-L^{12})y_t$

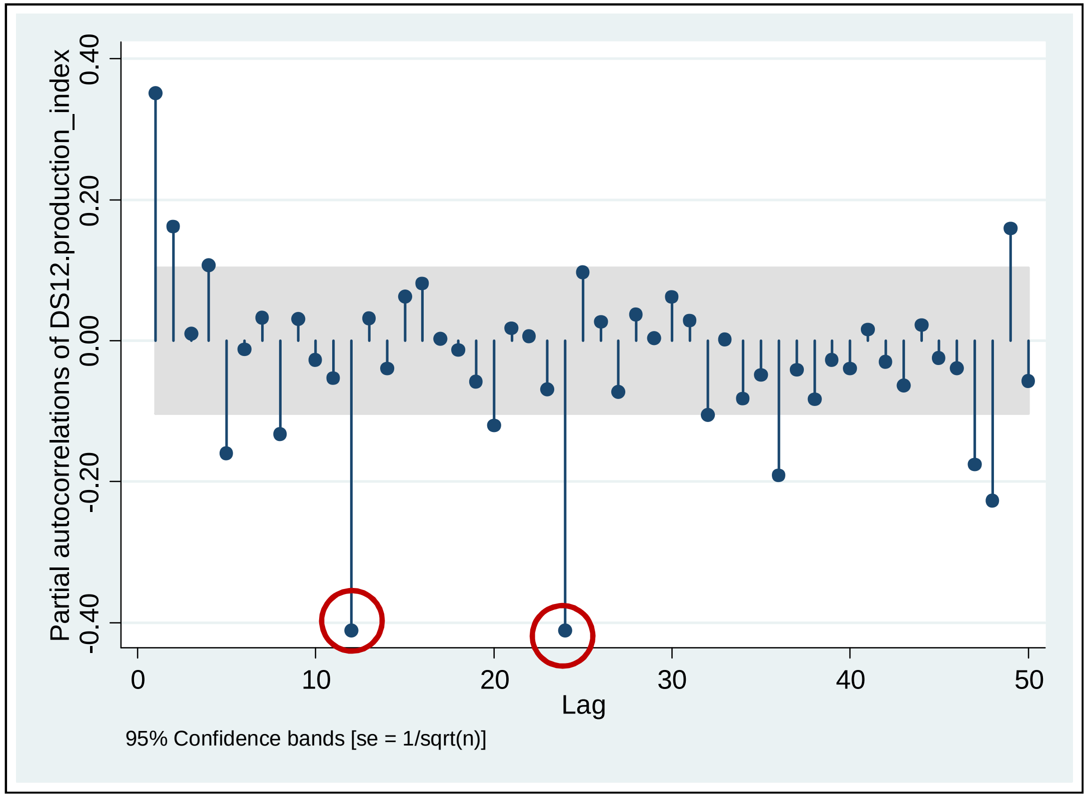


Characteristics of the ACF of this series tend to show a peak at 12.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

PACF of seasonal difference of the differenced series $(1-L)(1-L^{12})y_t$



Characteristics of the PACF of this series tend to show a peaks at 12, 24, …

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

Identification

- substantial peak in the PACF at lag $k = 1$, indicating a possible autoregressive series of order $p = 1$
- seasonal moving average of order $Q = 1$ or
- seasonal autoregression of possible order $P = 2$

→ **Two candidate models:**

- ARIMA $(1,1,0) \times (0,1,1)_{12}$
- ARIMA $(1,1,0) \times (2,1,0)_{12}$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Two candidate models:

- ARIMA (1,1,0) x $(0,1,1)_{12}$

$$(1-\phi_1 L)(1-L)(1-L^{12})y_t = (1-\theta_1^s L^{12})\varepsilon_t$$

$$(1-\phi_1 L)x_t = (1-\theta_1^s L^{12})\varepsilon_t$$

$$y_t = (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} - y_{t-12} - (1+\phi_1)y_{t-13} - \phi_1 y_{t-14} + \varepsilon_t + \theta_1^s \varepsilon_{t-12}$$

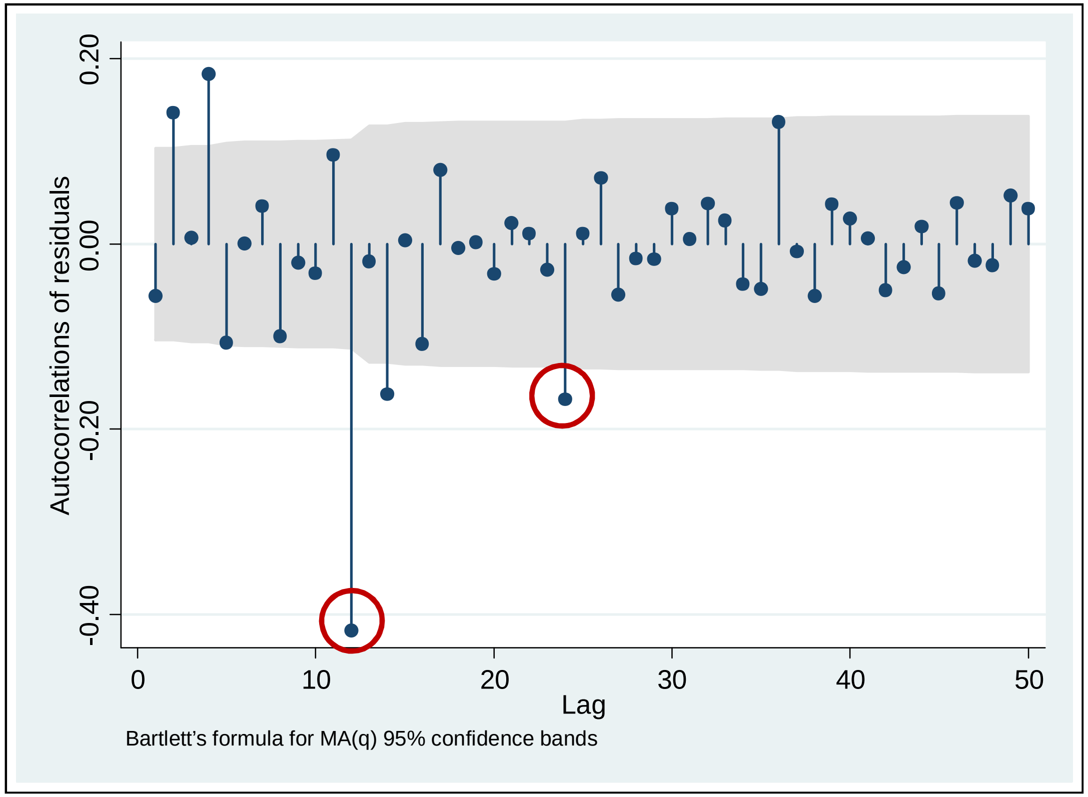
- ARIMA (1,1,0) x $(2,1,0)_{12}$

$$(1-\Phi_1^s L^{12} - \Phi_2^s L^{24})(1-\phi_1 L)(1-L)(1-L^{12})y_t = \varepsilon_t$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

ACF of residuals from the ARIMA (1,1,0) x $(0,1,0)_{12}$ model

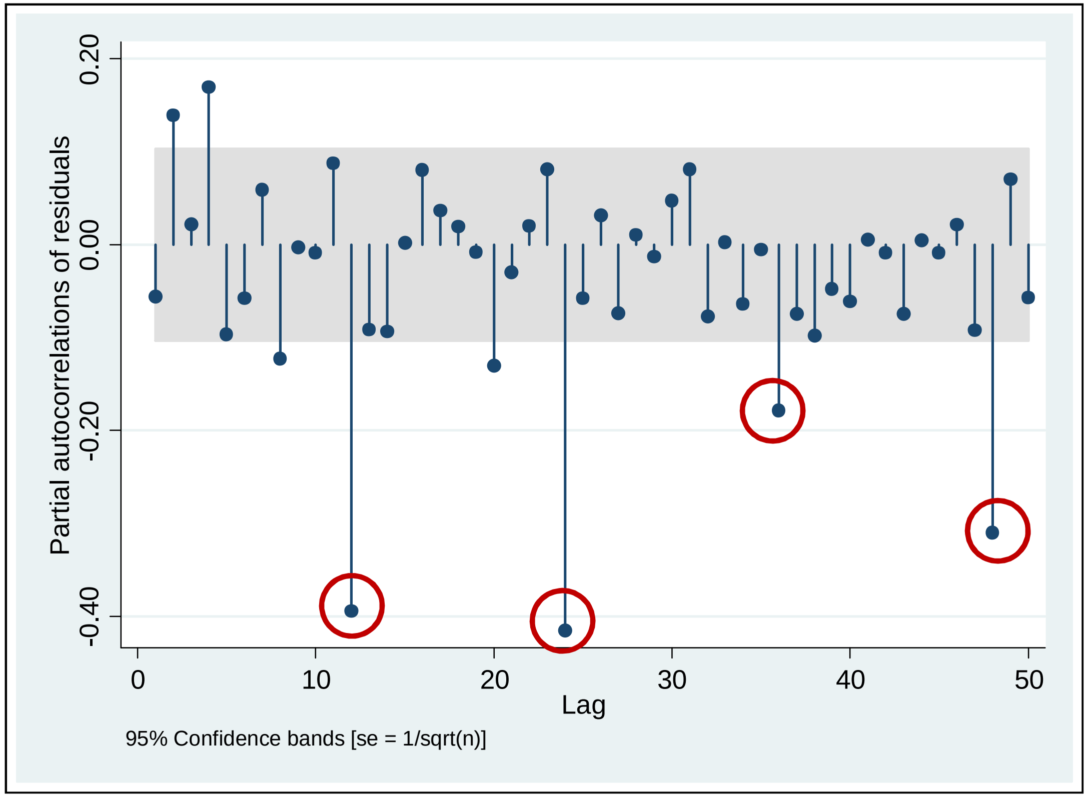


Continuing presence of seasonal peaks.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

PACF of residuals from the ARIMA (1,1,0) x $(0,1,0)_{12}$ model



Continuing presence of seasonal peaks.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

Estimation of ARIMA (1,1,0) x $(0,1,1)_{12}$

```
. arima production_index, arima(1,1,0) sarima(0,1,1,12) noconstant
[...]
ARIMA regression

Sample:  1949m2 to 1978m12                      Number of obs      =       359
                                                Wald chi2(2)       =    429.82
Log likelihood = -578.9588                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
DS12.        |
production~x |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3290424   .0278922    11.80   0.000     .2743747    .3837102
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.6896994   .0347539   -19.85   0.000    -.7578157    -.621583
-------------+----------------------------------------------------------------
      /sigma |   1.200597   .0269365    44.57   0.000     1.147803    1.253392
------------------------------------------------------------------------------
```

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Estimation of ARIMA (1,1,0) x $(2,1,0)_{12}$

```
. arima production_index, arima(1,1,0) sarima(2,1,0,12) noconstant
[...]
ARIMA regression

Sample:  1949m2 to 1978m12                      Number of obs      =       359
                                                Wald chi2(3)       =    383.71
Log likelihood = -582.0925                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
DS12.        |
production~x |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3605103   .0285088    12.65   0.000     .3046341    .4163866
-------------+----------------------------------------------------------------
ARMA12       |
          ar |
         L1. |  -.5984939   .0403742   -14.82   0.000    -.6776259    -.519362
         L2. |  -.4256483   .0405376   -10.50   0.000    -.5051004   -.3461961
-------------+----------------------------------------------------------------
      /sigma |   1.212154    .031385    38.62   0.000     1.150641    1.273668
------------------------------------------------------------------------------
```

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Diagnostic Checking

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q}{T}$$

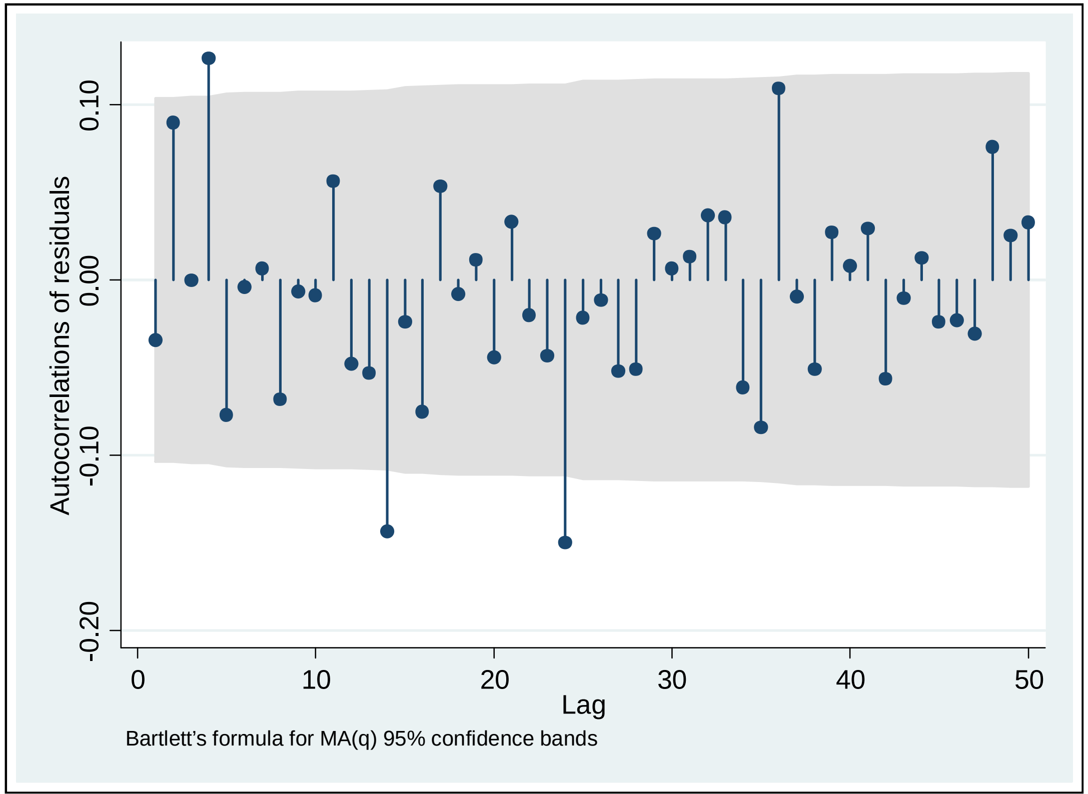
- AIC of model ARIMA (1,1,0) x $(0,1,1)_{12}$: 0.377
- AIC of model ARIMA (1,1,0) x $(2,1,0)_{12}$: 0.402
- AIC of the simpler model ARIMA (1,1,0) x $(2,1,0)_{12}$: 0.798

→ **So we tend to prefer the ARIMA (1,1,0) x $(0,1,1)_{12}$ model**

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

ACF of residuals from the ARIMA (1,1,0) x $(0,1,1)_{12}$ model

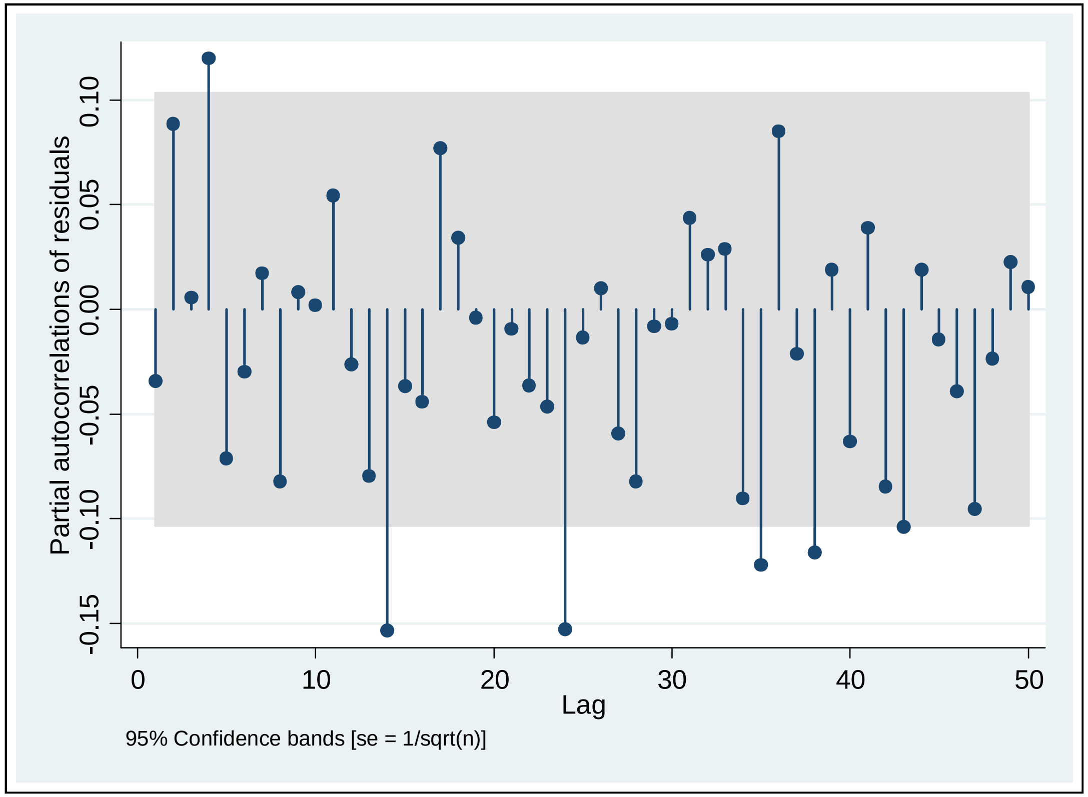


The ACF does not show patterns.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

PACF of residuals from the ARIMA (1,1,0) x $(0,1,1)_{12}$ model



The PACF does not show patterns.

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Diagnostic checking of ARIMA (1,1,0) x $(0,1,1)_{12}$

Joint Hypothesis Test

H_0: All autocorrelation coefficients are zero

Box and Ljung (refined test)

$$Q = T(T+2)\sum_{k=1}^{K}\frac{1}{T-K}\hat{\rho}_k^2 \quad \sim\chi^2$$

with $K-p-q$ degrees of freedom

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

Diagnostic checking of ARIMA (1,1,0) x $(0,1,1)_{12}$

```
. corrgram residuals, lags(24)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
[...]
12      -0.0476  -0.0261   15.125  0.2347          |                  |
[...]
24      -0.1498  -0.1529   38.138  0.0336         -|                 -|

. di 1-chi2(10, 15.125)
.12757061
. di 1-chi2(22, 38.138)
.01767923
```

- Possibility that there is still some slight seasonal regularity in the residuals.
- There are a few isolated outliers in the residuals, but other than these, the residuals behave as normal residuals.

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Estimation of ARIMA (1,1,0) x $(0,1,1)_{12}$

```
ARIMA regression

Sample:  1949m2 to 1978m12                      Number of obs      =       359
                                                Wald chi2(2)       =    429.82
Log likelihood = -578.9588                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |                 OPG
DS12.        |
production~x |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3290424   .0278922    11.80   0.000     .2743747    .3837102
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.6896994   .0347539   -19.85   0.000    -.7578157    -.621583
-------------+----------------------------------------------------------------
      /sigma |   1.200597   .0269365    44.57   0.000     1.147803    1.253392
------------------------------------------------------------------------------
```

$$\hat{\phi}_1 = 0.329 \qquad \hat{\theta}^{12} = -0.69 \qquad \hat{\sigma}_{\varepsilon}^2 = 1441$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Final Model: ARIMA (1,1,0) x $(0,1,1)_{12}$

$$(1-0.329L)(1-L^{12})(1-L)y_t = (1+0.69L^{12})\varepsilon_t$$

$$\begin{aligned}
&(1-0.329L)(1-L^{12})(1-L)y_t = (1+0.69L^{12})\varepsilon_t \\
&\Leftrightarrow (1-0.329L)(1-L^{12})(y_t - y_{t-1}) = \varepsilon_t + 0.69\varepsilon_{t-12} \\
&\Leftrightarrow (1-0.329L)(y_t - y_{t-1} - y_{t-12} + y_{t-13}) = \varepsilon_t + 0.69\varepsilon_{t-12} \\
&\Leftrightarrow y_t - y_{t-1} - y_{t-12} + y_{t-13} - 0.329y_{t-1} + 0.329y_{t-2} + 0.329y_{t-13} - 0.329y_{t-14} = \varepsilon_t + 0.69\varepsilon_{t-12} \\
&\Leftrightarrow y_t - 1329y_{t-1} + 0.329y_{t-2} - y_{t-12} + 1329y_{t-13} - 0.329y_{t-14} = \varepsilon_t + 0.69\varepsilon_{t-12} \\
&\Leftrightarrow y_t = 1329y_{t-1} - 0.329y_{t-2} + y_{t-12} - 1329y_{t-13} + 0.329y_{t-14} + \varepsilon_t + 0.69\varepsilon_{t-12}
\end{aligned}$$

$$\begin{aligned}
y_t = {} & 1329y_{t-1} - 0.329y_{t-2} + y_{t-12} - 1329y_{t-13} \\
& + 0.329y_{t-14} + \varepsilon_t + 0.69\varepsilon_{t-12}
\end{aligned}$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

ARMA(*p*,*q*) process at time *T + l*:

$$\tilde{y}_{T+l} = \varphi_1 \tilde{y}_{T+l-1} + \ldots + \varphi_p y_{T+l-p} + \tilde{\varepsilon}_{T+l} - \theta_1 \tilde{\varepsilon}_{T+l-1} - \ldots - \theta_q \tilde{\varepsilon}_{T+l-q|T}$$

Recursive forecasting recipe:

1. replace unknown y_{T+l} by their forecasts for $l > 0$;
2. "forecasts" of y_{T+l} , $l \leq 0$, are simply the known values y_{T+l}
3. since ε_t is white noise, the optimal forecast of ε_{T+l}, $l > 0$, is simply zero
4. "forecasts" of ε_{T+l}, $l \leq 0$, are just the known values ε_{T+l}

Example: Federal Reserve Board Production Index

Forecasting 12 month with ARIMA (1,1,0) x (0,1,1)$_{12}$

$$\hat{y}_{T+l|T} = 1329 y_{T+l-1|T} - 0.329 y_{T+l-2|T} + y_{T+l-12|T} - 1329 y_{T+l-13|T} + 0.329 y_{T+l-14|T} + \varepsilon_{T+l|T} - 0.69 \varepsilon_{T+l-12|T}$$

$$\hat{y}_{T+1|T} = 1329 y_{T+1-1|T} - 0.329 y_{T+1-2|T} + y_{T+1-12|T} - 1329 y_{T+1-13|T} + 0.329 y_{T+1-14|T} + \varepsilon_{T+1|T} - 0.69 \varepsilon_{T+1-12|T}$$

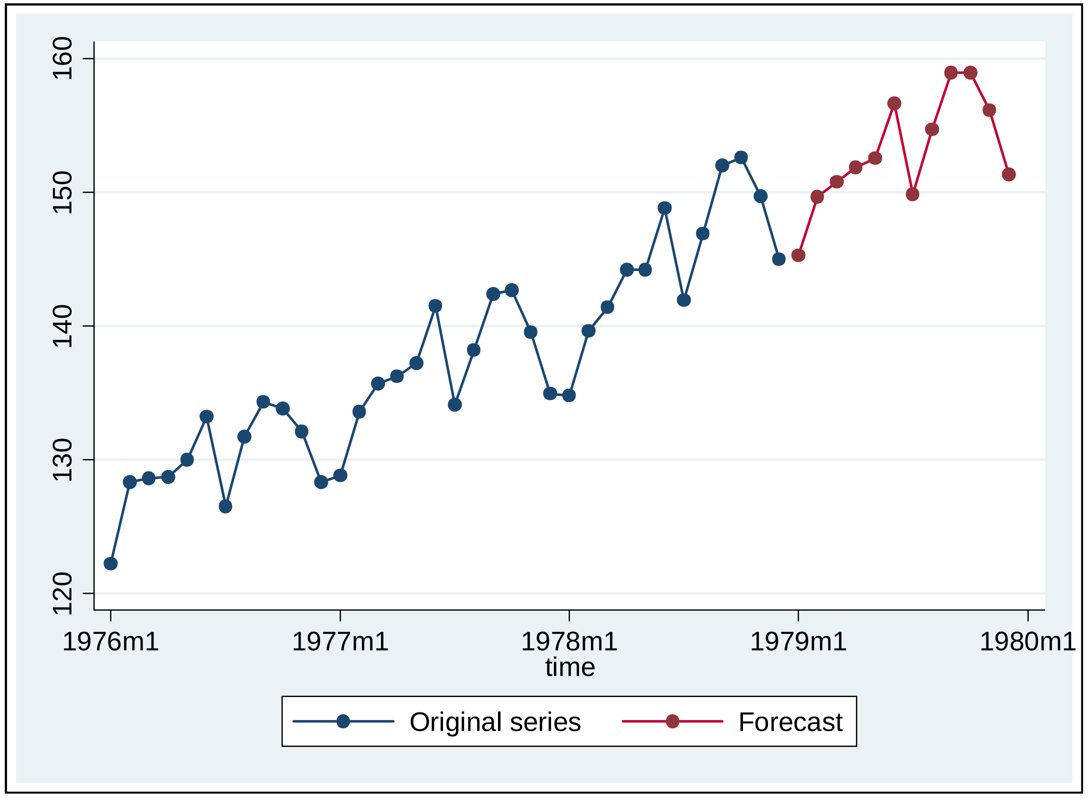
$$\hat{y}_{T+1|T} = 1329 y_{T|T} - 0.329 y_{T-1|T} + y_{T-11|T} - 1329 y_{T-12|T} + 0.329 y_{T-13|T} + \underbrace{\varepsilon_{T+1|T}}_{=0} - 0.69 \varepsilon_{T-11|T}$$

$$\begin{aligned} \hat{y}_{T+1|T} &= 1329 \cdot 145 - 0.329 \cdot 149.7 + 134.8 - 1329 \cdot 134.9 \\ &\quad + 0.329 \cdot 139.5 + 0.69 \cdot (-0.611) \\ &= 145.29 \end{aligned}$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Forecasting 12 month with ARIMA (1,1,0) x $(0,1,1)_{12}$



The future values are forecasted as a relatively linear extension to the current series that essentially repeats the last 12-month period.

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Recall: Forecast Error and Prediction Interval

General Solution "MA representation"

Forecast Error

$$MSE\left(\tilde{y}_{T+l|T}\right) = E\left[\left(y_{T+l} - \tilde{y}_{T+l|T}\right)^2\right] = \left(1 + \psi_1^2 + \ldots + \psi_{l-1}^2\right)\sigma_\varepsilon^2$$

Prediction Interval

$$y_{T+l} = \tilde{y}_{T+l|T} \pm 196\left(1 + \psi_1^2 + \ldots + \psi_{l-1}^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

$$= \tilde{y}_{T+l|T} \pm 196\left(\sum_{j=0}^{l-1}\psi_j^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

Harvey (1981), Time Series Models, p.157-164

How do we find $\psi_1, \ldots, \psi_{l-1}$?

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p\right) y_t = \left(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right) \varepsilon_t$$

$$a(L) y_t = b(L) \varepsilon_t$$

$$y_t = c(L) \varepsilon_t$$

$$= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with} \quad \psi_0 = 1$$

So the ψ_1, ψ_2, ... coefficients in $c(L)$, can be obtained by equating coefficients of L^j, $j = 1, 2, \ldots$ in $a(L)c(L) = b(L)$.

Example: Federal Reserve Board Production Index

Forecast errors ARIMA (1,1,0) x (0,1,1)$_{12}$

$$(1-0.329L)(1-L^{12})(1-L)(1+\hat{\psi}_1 L+\hat{\psi}_2 L^2+\ldots)=(1+0.69L^{12})$$

$$\begin{pmatrix} 1-L^{12}-L+L^{13}-0.329L+0.329L^{13} \\ +0.329L^2-0.329L^{14} \end{pmatrix}(1+\hat{\psi}_1 L+\hat{\psi}_2 L^2+\ldots)=(1+0.69L^{12})$$

$$L^1\!: -1-0.329+\hat{\psi}_1=0 \Rightarrow \hat{\psi}_1=1+0.329$$

$$L^2\!: 0.329-\hat{\psi}_1-0.329\hat{\psi}_1+\hat{\psi}_2=0 \Rightarrow \hat{\psi}_2=(1+0.329)\hat{\psi}_1-0.329$$

$$L^3\!: -\hat{\psi}_2-0.329\hat{\psi}_2+\hat{\psi}_3+0.329\hat{\psi}_1=0 \Rightarrow \hat{\psi}_3=(1+0.329)\hat{\psi}_2-0.329\hat{\psi}_1$$

$$\vdots$$

$$L^k\!: \Rightarrow \hat{\psi}_j=(1+0.329)\hat{\psi}_{j-1}-0.329\hat{\psi}_{j-2}$$

Shumway, Stoffer (2000) “Time Series analysis and Its Applications”, p. 161-166

Example: Federal Reserve Board Production Index

Forecast errors ARIMA (1,1,0) x (0,1,1)$_{12}$

$$\hat{y}_{T+l|T} \pm 196\left(\sum_{j=0}^{l-1}\psi_j^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

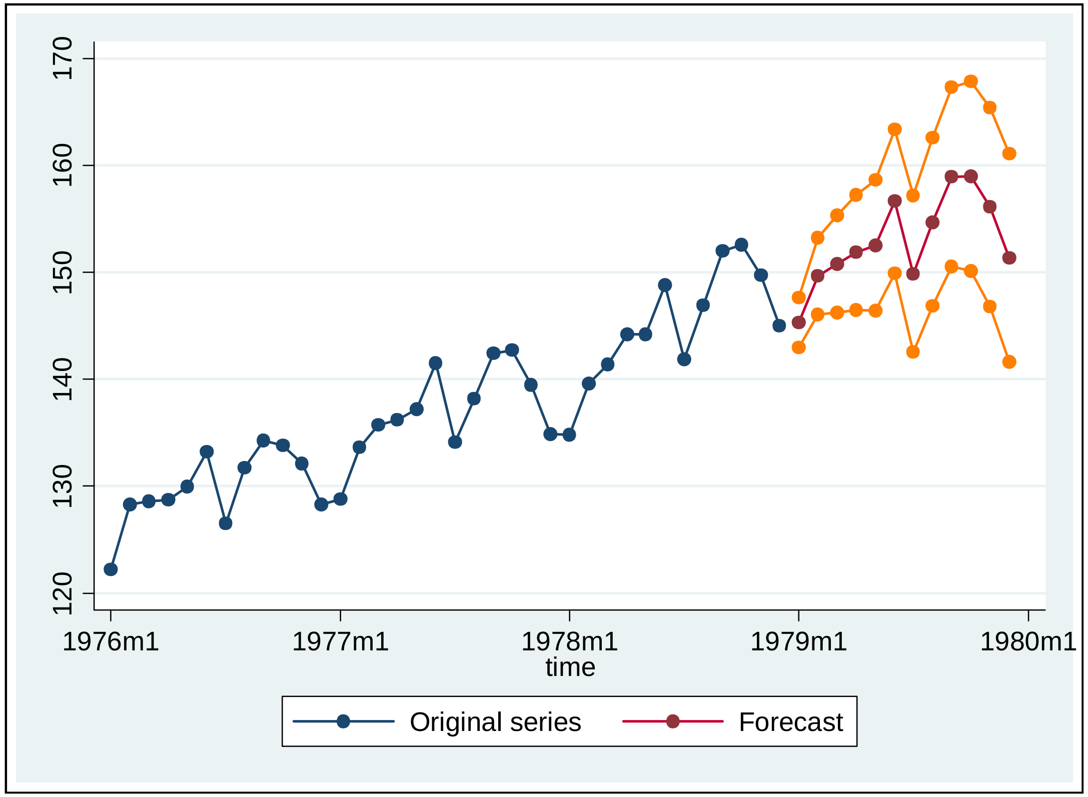
$$\hat{y}_{T+1|T} \pm 196\left(\sum_{j=0}^{1-1}\hat{\psi}_j^2\right)^{\frac{1}{2}}\hat{\sigma}_\varepsilon = 145.29 \pm 196 \cdot (1)^{\frac{1}{2}} \cdot 12$$

$$= 145.29 \pm 2.352$$

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166

Example: Federal Reserve Board Production Index

Prediction Interval of 12 month forecast with ARIMA (1,1,0) x $(0,1,1)_{12}$



The 95% upper and lower prediction limits are broad, as it is customary for most forecasts.

Because the difference operator has roots on the unit circle, the representation for the estimation error of the finite approximation to the forecast based on the infinite past will be poor.

Shumway, Stoffer (2000) "Time Series analysis and Its Applications", p. 161-166