universität
innsbruck

Time Series Analysis

Models for Nonstationary Time Series

Models for Nonstationary Time Series

Contents

Contents

Models for Nonstationary Time Series

Stationarity through Differencing

Stationarity through differencing

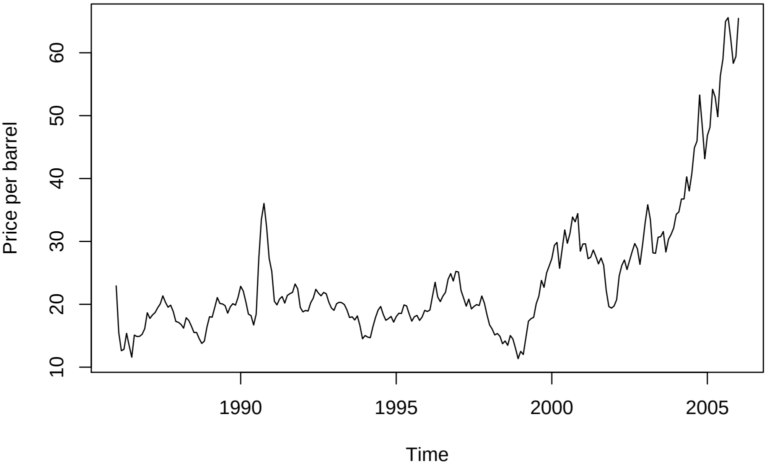
Motivation: Any series with nonconstant mean is nonstationary.

Goal: Model nonstationarity.

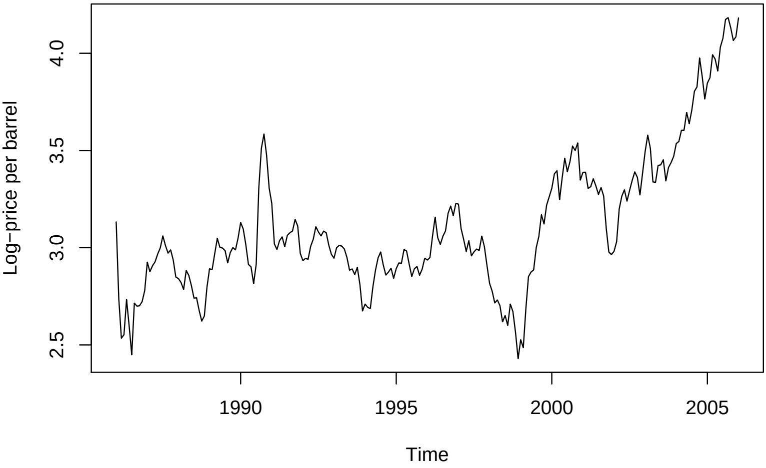
First approach: *Trend stationarity*, i.e., $Y_t = \mu_t + X_t$ with deterministic μ_t and stationary X_t (see Chapter 3). Easy to estimate *but* requires stable trend that continues “forever”.

Second approach: *Difference stationarity.* Process is stationary after differencing. More reasonable for many time series in business and economics.

Stationarity through differencing



Stationarity through differencing



Stationarity through differencing

Example: Consider again AR(1) model

$$Y_t = \phi Y_{t-1} + e_t$$

with innovations e_t (i.e., uncorrelated with $Y_{t-1}, Y_{t-2}, \ldots$). For a stationary AR(1) process, we must have $|\phi| < 1$.

Question: What are the properties for $|\phi| > 1$?

Illustration: For $\phi = 3$

$$\begin{aligned} Y_t &= 3Y_{t-1} + e_t \\ &= e_t + 3e_{t-1} + 3^2 e_{t-2} + \ldots + 3^{t-1} e_1 + 3^t Y_0 \end{aligned}$$

Thus: The influence of past values does not die out – it grows exponentially large.

Stationarity through differencing

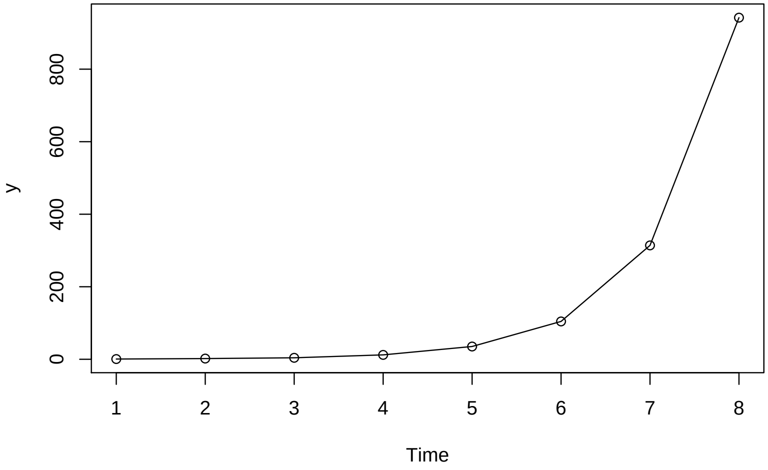
In R: Using standard normal white noise and $Y_0 = 0$.

```
R> set.seed(44)
R> e <- ts(rnorm(8))
R> y <- filter(e, 3, method = "recursive")
R> ts.intersect(e, y)
```

```
Time Series:
Start = 1
End = 8
Frequency = 1
         e        y
1  0.65392   0.6539
2  0.01905   1.9808
3 -1.84950   4.0929
4 -0.13276  12.1460
5 -1.19882  35.2391
6 -1.32974 104.3877
7  0.91649 314.0796
8 -0.16295 942.0758
```

```
R> plot(y, type = "o")
```

Stationarity through differencing



Stationarity through differencing

Interpretation: Explosive behavior.

$$
\begin{aligned}
\mathrm{Var}(Y_t) &= \frac{1}{8}\left(9^t - 1\right)\sigma_e^2 \\
\mathrm{Cov}(Y_t, Y_{t-k}) &= \frac{3^k}{8}\left(9^{t-k} - 1\right)\sigma_e^2 \\
\mathrm{Cor}(Y_t, Y_{t-k}) &= 3^k\sqrt{\frac{9^{t-k} - 1}{9^t - 1}}
\end{aligned}
$$

The correlation is approximately 1 for large t and moderate k.

Analogously: Similar exponential growth for all other $|\phi| > 1$.

Stationarity through differencing

More reasonable: $\phi = 1$, yielding random walk.

$$Y_t = Y_{t-1} + e_t$$

Alternatively: Use *first differences* notation

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= e_t \end{aligned}$$

Idea: First differences can be some other stationary process, not just white noise.

Notation: Sometimes ∇ is used instead of Δ as the difference operator.

Stationarity through differencing

Suppose:

$$Y_t = M_t + X_t$$

where M_t is a slowly varying mean.

Assuming M_t is approximately constant for every two consecutive time points, it can be estimated via

$$\hat{M}_t = \frac{1}{2}\left(Y_t + Y_{t-1}\right)$$

The detrended series is then

$$\begin{aligned} Y_t - \hat{M}_t &= Y_t - \frac{1}{2}\left(Y_t + Y_{t-1}\right) \\ &= \frac{1}{2}\left(Y_t - Y_{t-1}\right) \\ &= \frac{1}{2}\,\Delta Y_t \end{aligned}$$

Stationarity through differencing

Alternatively: Assume that M_t is a random walk.

$$\begin{aligned} Y_t &= M_t + e_t \\ M_t &= M_{t-1} + \varepsilon_t \end{aligned}$$

where $\{e_t\}$ and $\{\varepsilon_t\}$ are independent white noise processes.

Then:

$$\begin{aligned} \Delta Y_t &= \Delta M_t + \Delta e_t \\ &= \varepsilon_t + e_t - e_{t-1} \end{aligned}$$

which has MA(1) autocorrelation with

$$\varrho_1 = -\frac{1}{2 + \sigma_\varepsilon^2 / \sigma_e^2}$$

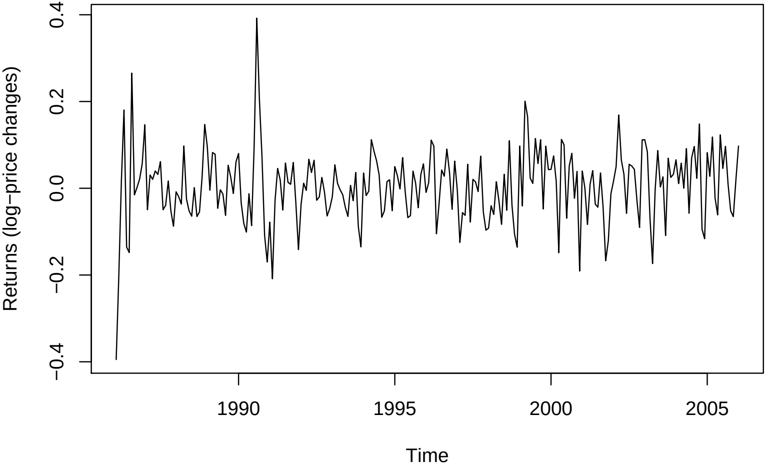
Stationarity through differencing

Conclusion: In all these situations ΔY_t can be studied as a stationary process.

Illustration: First differences of log-prices for oil series.

Economic interpretation: $\Delta \log(Y_t)$ are the (log-)returns pertaining to price series Y_t.

Stationarity through differencing



Stationarity through differencing

Furthermore: Assumptions can be made that lead to stationary second-difference models.

Example: Assuming mean M_t of a series Y_t is approximately linear in time over three consecutive time points, it can be estimated via

$$\hat{M}_t = \frac{1}{3}\left(Y_{t+1} + Y_t + Y_{t-1}\right)$$

The detrended series is then

$$\begin{aligned}
Y_t - \hat{M}_t &= Y_t - \frac{1}{3}\left(Y_{t+1} + Y_t + Y_{t-1}\right) \\
&= -\frac{1}{3}\left(Y_{t+1} - 2Y_t + Y_{t-1}\right) \\
&= -\frac{1}{3}\left[\left(Y_{t+1} - Y_t\right) - \left(Y_t - Y_{t-1}\right)\right] \\
&= -\frac{1}{3}\,\Delta(\Delta Y_{t+1}) = -\frac{1}{3}\,\Delta^2 Y_{t+1}
\end{aligned}$$

Stationarity through differencing

Similarly: Stochastic trend M_t whose "rate of change" ΔM_t is changing only slowly.

$$\begin{aligned} Y_t &= M_t + e_t \\ M_t &= M_{t-1} + W_t \\ W_t &= W_{t-1} + \varepsilon_t \end{aligned}$$

Then:

$$\begin{aligned} \Delta Y_t &= \Delta M_t + \Delta e_t \\ &= W_t + \Delta e_t \\ \Delta^2 Y_t &= \Delta W_t + \Delta^2 e_t \\ &= \varepsilon_t + (e_t - e_{t-1}) - (e_{t-1} - e_{t-2}) \\ &= \varepsilon_t + e_t - 2e_{t-1} + e_{t-2} \end{aligned}$$

which has the autocorrelation function of an MA(2) process.

Stationarity through differencing

Remarks:

- $\Delta^2 Y_t$ is the *second difference* of Y_t at *lag 1*.
- Main idea: Differences of nonstationary series can be stationary.
- Differencing often has economic interpretation.
- Nonstationary series that need to be differenced twice (or more) are rare in practice. (Some authors argue that certain monetary quantities need double differencing.)

Models for Nonstationary Time Series

ARIMA Models

ARIMA models

Definition: $\{Y_t\}$ is an *integrated autoregressive moving average process of order (p, d, q)*, ARIMA(p, d, q) for short, if $\{W_t\}$ with $W_t = \Delta^d Y_t$ is a stationary ARMA(p, q) process.

$$\begin{aligned} W_t &= \Delta^d Y_t \\ W_t &= \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}. \end{aligned}$$

Remarks:

- Without autoregressive terms – ARIMA(0, d, q) – it is sometimes also called *integrated moving average* process IMA(d, q).
- Without moving average terms – ARIMA(p, d, 0) – it is sometimes also called *integrated autoregressive* process ARI(p, d).

ARIMA models: ARIMA(p, 1, q)

Example: ARIMA(p, 1, q).

Rewrite in terms of observed series

$$\begin{aligned} Y_t - Y_{t-1} \;&=\; \phi_1(Y_{t-1} - Y_{t-2}) \,+\, \ldots \,+\, \phi_p(Y_{t-p} - Y_{t-p-1}) \,+ \\ &\quad\; e_t \,-\, \theta_1 e_{t-1} \,-\, \ldots \,-\, \theta_q e_{t-q}. \end{aligned}$$

Or in so-called *difference equation form*

$$\begin{aligned} Y_t \;&=\; (1+\phi_1)Y_{t-1} \,+\, (\phi_2 - \phi_1)Y_{t-2} \,+\, \ldots \,+\, (\phi_p - \phi_{p-1})Y_{t-p} \,- \\ &\quad\; \phi_p Y_{t-p-1} \,+\, e_t \,-\, \theta_1 e_{t-1} \,-\, \ldots \,-\, \theta_q e_{t-q}. \end{aligned}$$

Note that this appears to be an ARMA($p+1$, q) process.
However, it is not stationary.

ARIMA models: ARIMA(p, 1, q)

Characteristic polynomial

$$\begin{aligned} 1 \;-\; (1+\phi_1)x \;-\; (\phi_2-\phi_1)x^2 \;-\; \dots \;-\; (\phi_p-\phi_{p-1})x^p \;+\; \phi_p x^{p+1} \\ =\; \left(1-\phi_1 x-\phi_2 x^2-\cdots-\phi_p x^p\right)(1-x) \end{aligned}$$

Thus, there is a unit root. However, all remaining roots are those of the stationary process ΔY_t.

Explicit representation:

- As not in equilibrium, Y_t cannot start in $t = -\infty$.
- However, a starting point $t = -m$ (before first observation at $t = 1$) can be assumed.
- For convenience $Y_t = 0$ for $t < -m$.
- Then

$$Y_t \;=\; \sum_{j=-m}^{t} W_j.$$

ARIMA models: ARIMA(p, 2, q)

Analogously: ARIMA(p, 2, q).

Explicit representation:

$$
\begin{aligned}
Y_t &= \sum_{j=-m}^{t} \sum_{i=-m}^{j} W_i \\
&= \sum_{j=0}^{t+m} (j+1)\, W_{t-j}
\end{aligned}
$$

Remark: Explicit representation have limited practical use but can help the theoretic understanding (e.g., expectation, covariance, etc.).

ARIMA models: ARIMA(0, 1, 1)

Example: ARIMA(0, 1, 1). Simple model that satisfactorily represents many time series in business and economics.

In difference equation form:

$$\begin{aligned} Y_t &= Y_{t-1} + e_t - \theta e_{t-1} \\ &= e_t + (1-\theta)e_{t-1} + \ldots + (1-\theta)e_{-m} - \theta e_{-m-1} \end{aligned}$$

Remarks:

- Weights do not die out.
- Y_t is an almost equally weighted accumulation of many white noise values.

ARIMA models: ARIMA(0, 1, 1)

Variance and covariance function:

$$\begin{aligned}
\mathrm{Var}(Y_t) &= \left[1+\theta^2+(1-\theta)^2(t+m)\right]\sigma_e^2 \\
\mathrm{Cor}(Y_t, Y_{t-k}) &= \frac{\left[1+\theta^2+(1-\theta)^2(t+m-k)-2\theta\right]\ \sigma_e^2}{\sqrt{\mathrm{Var}(Y_t)\ \mathrm{Var}(Y_{t-k})}} \\
&\approx \sqrt{\frac{t+m-k}{t+m}}
\end{aligned}$$

Thus:

- $\mathrm{Var}(Y_t)$ increases with t and can be vary large.
- $\mathrm{Cor}(Y_t, Y_{t-k})$ is strongly positive for many lags k. It is close to 1 for large t and moderate k.

ARIMA models: ARIMA(0, 2, 2)

Example: ARIMA(0, 2, 2).

In difference equation form:

$$\begin{aligned}
\Delta^2 Y_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\
Y_t &= 2Y_{t-1} - Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\
&= e_t + \sum_{j=1}^{t+m} \left[1 + \theta_2 + (1 - \theta_1 - \theta_2)\, j\right] e_{t-j} - \\
&\quad [(t+m+1)\theta_1 + (t+m)\theta_2] e_{-m-1} - (t+m+1)\theta_2 e_{-m-2}
\end{aligned}$$

Similarly:

- $\text{Var}(Y_t)$ increases rapidly with t.
- $\text{Cor}(Y_t, Y_{t-k})$ is nearly 1 for all moderate k.

ARIMA models: ARIMA(0, 2, 2)

In R: Simulation with `arima.sim()`.

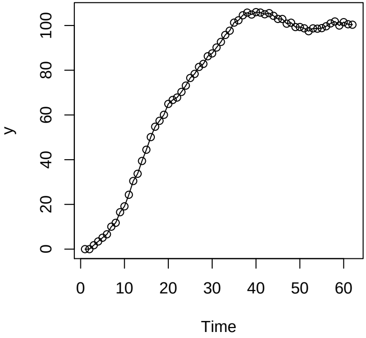
Here: $\theta_1 = 1$, $\theta_2 = -0.6$.

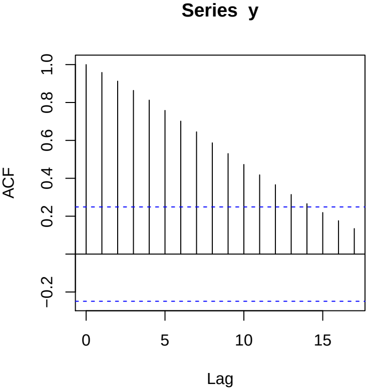
```
R> set.seed(42)
R> y <- arima.sim(list(order = c(0, 2, 2), ma = c(-1, 0.6)), n = 60)
R> plot(y, type = "o")
R> plot(diff(y), type = "o")
R> plot(diff(y, differences = 2), type = "o")
```

Remarks:

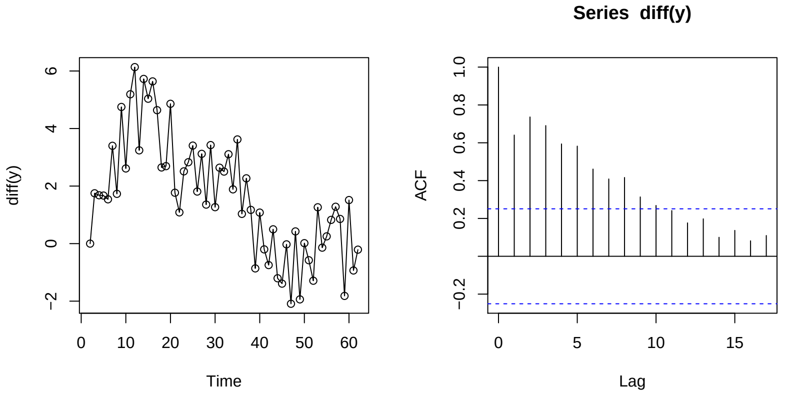
- Process changes very smoothly.
- Zero mean unimportant in single trajectory.

ARIMA models: ARIMA(0, 2, 2)

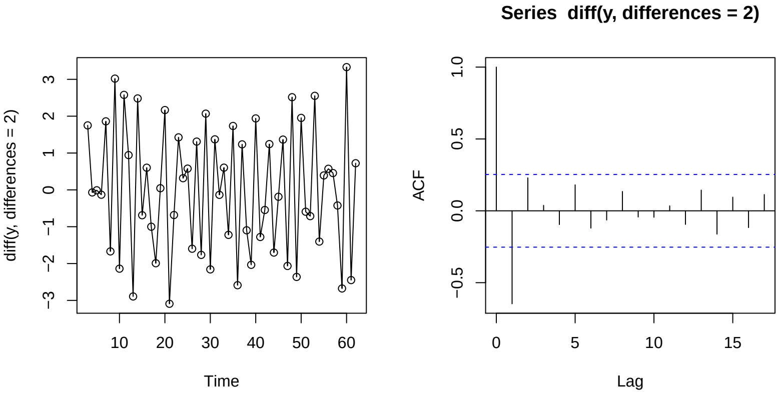




ARIMA models: ARIMA(0, 2, 2)



ARIMA models: ARIMA(0, 2, 2)



Models for Nonstationary Time Series

Constant Terms in ARIMA Models

Constant terms in ARIMA models

Up to now: Assumed zero mean for stationary processes.

Question: How can a nonzero constant mean be included in an ARMA (component of an ARIMA) process?

Answers: Add mean/constant for $W_t = \Delta^d Y_t$.

1. ARMA process holds for $W_t - \mu$.
2. W_t follows ARMA process plus constant θ_0.

Both approaches lead equivalent models.

Constant terms in ARIMA models

First approach yields

$$\begin{aligned} W_t - \mu &= \phi_1(W_{t-1} - \mu) + \ldots + \phi_p(W_{t-p} - \mu) + \\ &\quad e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \end{aligned}$$

Second approach yields

$$\begin{aligned} W_t &= \theta_0 + \phi_1 W_{t-1} + \ldots + \phi_p W_{t-p} + \\ &\quad e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \end{aligned}$$

Taking expectations yields

$$\begin{aligned} \mu &= \theta_0 + (\phi_1 + \cdots + \phi_p)\,\mu \\ &= \frac{\theta_0}{1 - \phi_1 - \cdots - \phi_p} \\ \theta_0 &= \mu\,(1 - \phi_1 - \cdots - \phi_p) \end{aligned}$$

Constant terms in ARIMA models

Question: What is the effect on the undifferenced series Y_t?

Example: ARIMA(0, 1, 1) with constant term.

$$\begin{aligned} Y_t &= Y_{t-1} + \theta_0 + e_t - \theta e_{t-1} \\ &= e_t + (1-\theta)e_{t-1} + \ldots + (1-\theta)e_{-m} - \theta e_{-m-1} + \\ &\quad (t+m+1)\theta_0 \end{aligned}$$

Interpretation: Adding constant to W_t accumulates to additional linear trend (or drift) with slope θ_0 in Y_t.

Equivalently: Y_t can be represented as

$$Y_t = Y_t' + \beta_0 + \beta_1 t$$

where Y_t' is ARIMA(0, 1, 1) with $\mathrm{E}(\Delta Y_t') = 0$ and $\mathrm{E}(\Delta Y_t) = \beta_1$.

Constant terms in ARIMA models

More generally: ARIMA(p, d, q) with $\mathrm{E}(\Delta^d Y_t) \neq 0$.

This can be written as

$$Y_t \;=\; Y'_t \;+\; \mu_t$$

with

- μ_t is a deterministic polynomial of degree d.
- Y'_t is ARIMA(p, d, q) with $\mathrm{E}(\Delta^d Y'_t) = 0$.

Models for Nonstationary Time Series

Other Transformations

Other transformations: Logarithms

Motivation: Dispersion often increases along with level of series.

Specifically

$$
\begin{aligned}
\mathrm{E}(Y_t) &= \mu_t & \sqrt{\mathrm{Var}(Y_t)} &= \mu_t\,\sigma \\
\mathrm{E}(\log(Y_t)) &\approx \log(\mu_t) & \mathrm{Var}(\log(Y_t)) &= \sigma^2
\end{aligned}
$$

due to first order Taylor series expansion

$$
\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}
$$

Thus:

- Proportional standard deviations are transformed to approximately constant variances.
- Exponential trends are transformed to roughly linear trends.

Other transformations: Percentage changes

Motivation: Stable percentage changes from one period to the next.

With percentage changes of $100X_t$ with stationary X_t

$$Y_t \; = \; (1 + X_t)\, Y_{t-1}$$

Then

$$\begin{aligned}
\log(Y_t) \; - \; \log(Y_{t-1}) \quad &= \quad \log\left(\frac{Y_t}{Y_{t-1}}\right) \\
&= \quad \log(1 + X_t) \\
&\approx \quad X_t
\end{aligned}$$

where the latter holds roughly for $|X_t| < 0.2$, i.e., percentage changes of at most $\pm 20\%$.

Other transformations: Returns

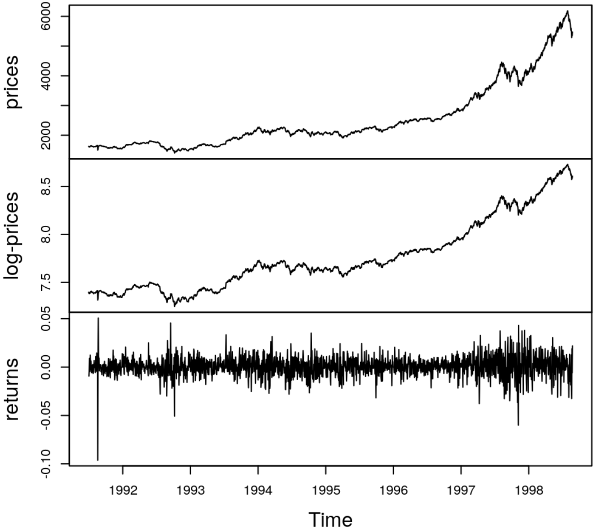
Hence: For many (economic) series consider *returns* $\Delta[\log(Y_t)]$.

Remarks:

- These are the returns in continuous time.
- Also called log-returns.
- Returns in discrete time $(Y_t - Y_{t-1})/Y_{t-1}$ are also called arithmetic returns.
- Can often be modeled by a stationary process.
- Differencing and taking logs may not be interchanged.

Illustration: DAX closing prices, log-prices, and returns.

Other transformations: Returns



Other transformations: Power transformations

Motivation: Generalization of log-transformation.

Power transformation or *Box-Cox transformation*

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log(x) & \text{for } \lambda = 0 \end{cases}$$

Remarks:

- Special cases: square root $\lambda = 1/2$ and reciprocal transformation $\lambda = -1$.
- Value of λ often determined subjectively, sometimes via using maximum likelihood (ML) assuming normal distribution for transformed data.

Notation: Backshift operator

Goal: Condense notation when working with lags.

Definition: The *backshift operator*, denoted B, operates on the time index and shifts time back one unit to form a new series.

$$BY_t = Y_{t-1}$$

Alternatively, this is sometimes denoted L and called *lag operator*.

Operator is linear

$$B(aY_t + bX_t + c) = aBY_t + bBX_t + c$$

Arithmetic almost as usual:

$$\begin{aligned} Y_{t-2} &= BY_{t-1} = BBY_t = B^2Y_t \\ Y_{t-m} &= B^mY_t \end{aligned}$$

Notation: Backshift operator

Rewrite: MA(q) model.

$$\begin{aligned} Y_t &= e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \\ &= e_t - \theta_1 B e_t - \ldots - \theta_q B^q e_t \\ &= (1 - \theta_1 B - \cdots - \theta_q B^q) e_t \\ &= \theta(B) e_t \end{aligned}$$

where $\theta(B)$ is the MA characteristic polynomial "evaluated" at B.

Analogously: AR(p) model.

$$\begin{aligned} e_t &= Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} \\ &= (1 - \phi_1 B - \cdots - \phi_p B^p) Y_t \\ &= \phi(B) Y_t \end{aligned}$$

where $\phi(B)$ is the AR characteristic polynomial "evaluated" at B.

Notation: Backshift operator

Rewrite: ARMA(p, q)

$$\begin{aligned} \phi(B)Y_t &= \theta(B)e_t \\ Y_t &= \psi(B)e_t \end{aligned}$$

where $\psi(x) = \theta(x)/\phi(x)$ is the polynomial for the general linear process form.

Moreover: First differences.

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= Y_t - BY_t \\ &= (1 - B)Y_t \end{aligned}$$

i.e., $\Delta = 1 - B$ and $\Delta^d = (1 - B)^d$.

Notation: Backshift operator

Analogously: ARIMA(p, d, q)

$$\phi(B)(1 - B)^d Y_t \; = \; \theta(B) e_t$$

Remarks:

- Notation employed in large parts of literature.
- Some care required for distinguishing B as backshift operator and B as real (or complex) variable.
- Example for use of B as auxiliary variable: “Roots of $\phi(B) = 0$ must lie outside unit circle (or equivalently be greater than 1 in absolute value).”