Seasonality

Matthieu Stigler
Matthieu.Stigler@gmail.com

November 14, 2008

Version 1.1

Lectures list

1. Stationarity
2. ARMA models for stationary variables
3. **Seasonality**
4. Non-stationarity
5. Non-linearities
6. Multivariate models
7. Structural VAR models
8. Cointegration the Engle and Granger approach
9. Cointegration 2: The Johansen Methodology
10. Multivariate Nonlinearities in VAR models
11. Multivariate Nonlinearities in VECM models

Outline

Decomposition

We have the traditional idea

$$Y_t = S_t + C_t + T_t + \varepsilon_t$$

- S_t: Seasonal component
- C_t: Cyclical component
- T_t: Trend component
- ε_t: Stochastic/unexplained component

Types of seasonality

Remember the last discussion:

- Stationarity: ARMA(p,q)
- Non-stationarity:
  - Trend stationarity (deterministic trend)
  - Difference stationarity (stochastic trend) ARIMA(p,d,q)

Types of seasonality

Remember the last discussion:

- Stationarity: ARMA(p,q)
- Non-stationarity:
  - Trend stationarity (deterministic trend)
  - Difference stationarity (stochastic trend) ARIMA(p,d,q)

Similarly we will have:

- Stationarity: ARMA$(p, q)(P, Q)_s$
- Non-stationarity:
  - Stochastic seasonality: ARIMA $(p, d, q)(P, D, Q)_s$
  - Deterministic seasonality: dummy variables

Outline

Dummy variable

Definition (Dummy Variable)

$$D = \begin{cases} 1 & \text{if } A \\ 0 & \text{if } \bar{A} \end{cases}$$

A can be:

- Male (and then $\bar{A}$ = female)
- Special event
- ...

Deterministic seasonality

Adds dummy variables:

$$y_t = \alpha_0 + \sum_{i=1}^{s-1} \alpha_i D_i + \beta t + \varepsilon_t$$

Adds sines and cosines:

$$y_t = \alpha_0 + \sum_{i=1}^{[T/2]} (\alpha_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)) + \varepsilon_t$$

With $\lambda_i = \frac{2\pi i}{T}$ $\qquad i = 1, \ldots, [T/2]$

Outline

Outline

Seasonal AR models

See:

$$x_t = a_4 x_{t-4} + \varepsilon_t$$

Autocorrelation: spikes at lags 1, 1s, 2s,

$$\gamma(j) = \begin{cases} a_4^{j/4} & \text{if } j/4 \in \mathcal{N} \\ 0 & \text{if } j/4 \notin \mathcal{N} \end{cases}$$

Define the pure AR seasonal model:

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \ldots + \Phi_P Y_{t-Ps} + \varepsilon_t$$

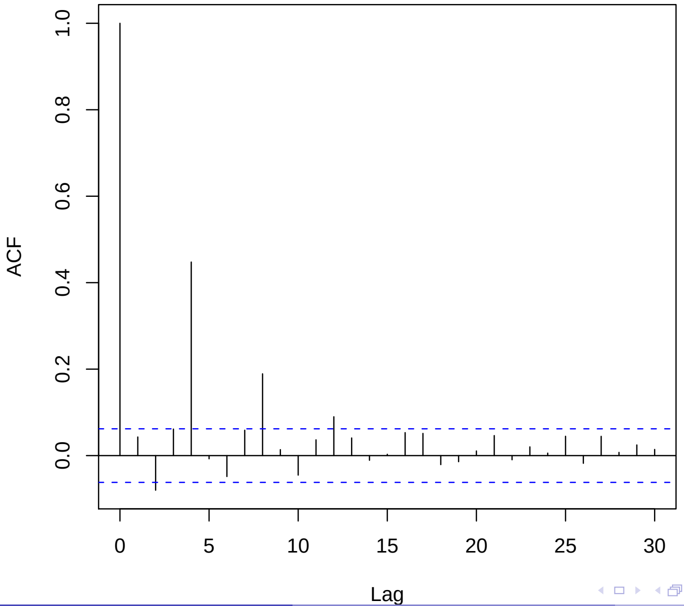
With polynomial:

$$\Phi^s(x) = (1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps})$$

Proposition

It can be shown that the autocorrelation function is nonzero only at lags s, 2s, 3s

Pure seasonal model, with a4=0.5 and others=0



Extending the ARMA model

Multiplicative way: ARMA$(p, q)(P, Q)_s$

$$\Phi(L^s)\phi(L)y_t = \Theta(L^s)\theta(L)\varepsilon_t$$

Multiplicative way: ARIMA$(p, d, q)(P, D, Q)_s$

$$\Phi(L^s)\phi(L)\Delta^d\Delta_s^D y_t = \Theta(L^s)\theta(L)\varepsilon_t$$

Outline

Seasonal unit roots

Take the simple seasonal AR(4) model:

$$x_t = a_4 x_{t-4} + \varepsilon_t$$

With $a_4 = 1$

Its polynomial is $(1 - L^4)$ and can be decomposed as:

$$(1 - L)(1 + L)(1 + L^2) = (1 - L)(1 + L)(1 - iL)(1 + iL)$$

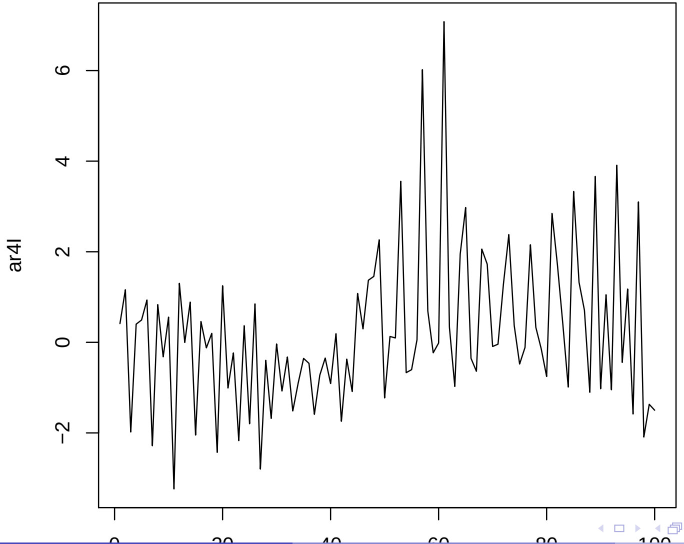
It has the four roots:

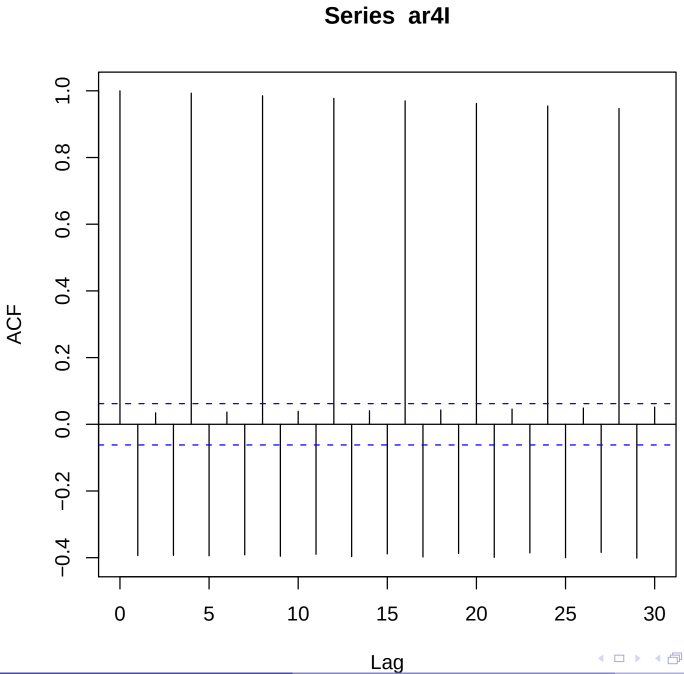
- 1
- -1
- i
- -i

So it is not stationary:

- Long/infinite memory
- variance increasing with the time

seasonal AR process with $\phi_4 = 1$




Series ar4I
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
5
10
15
20
25
30

Seasonal differencing

remember:

Definition (Difference operator)

$\Delta^d = (1 - L)^d$

Now:

Definition (Seasonal difference operator)

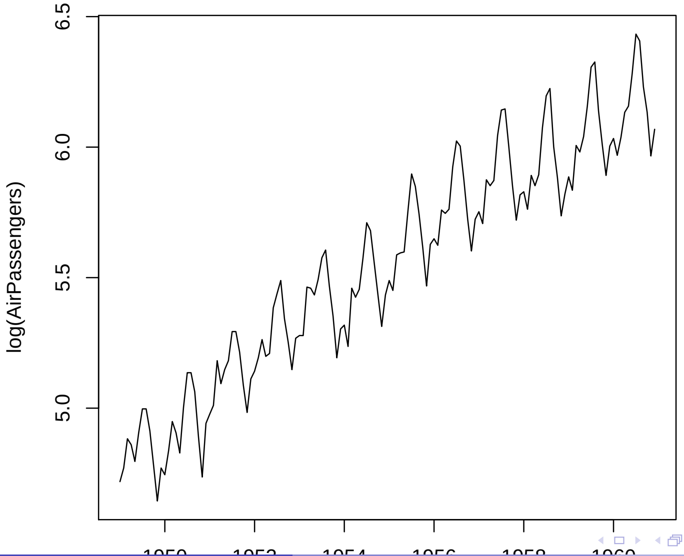
$\Delta_s^D = (1 - L^s)^D$

Definition (Seasonal integration)

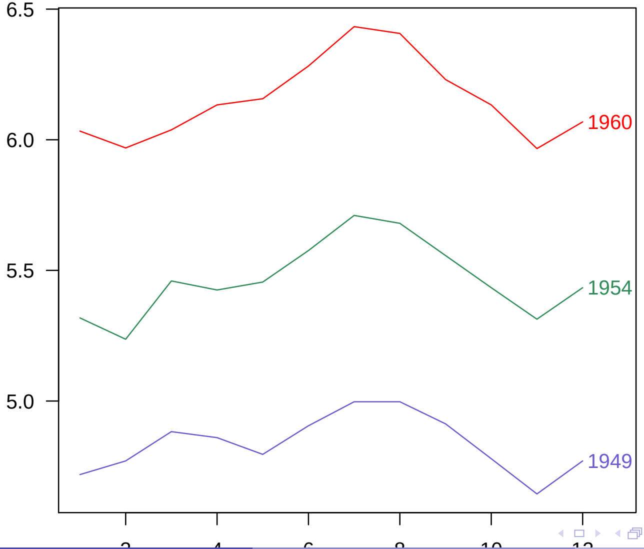
A series is said to be seasonaly integrated of order D if $\Delta_D y_t$ is stationary.

Outline

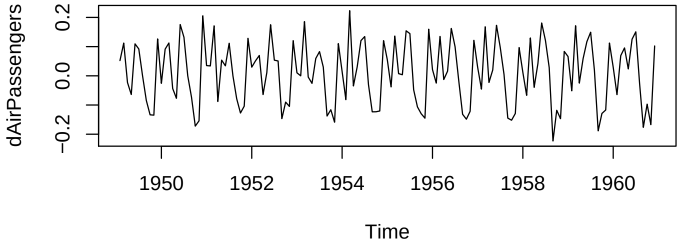
Box-Jenkins airline series (in log)

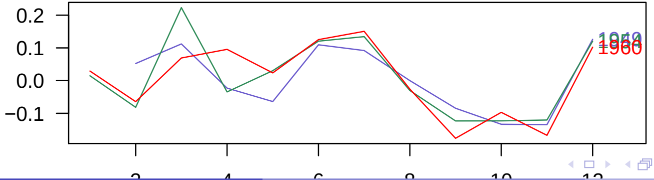


Box-Jenkins airline model

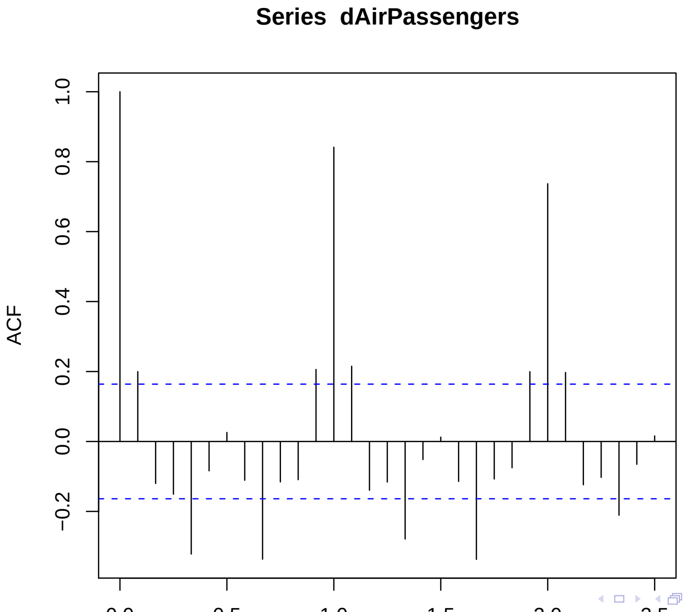


airline series: Δ

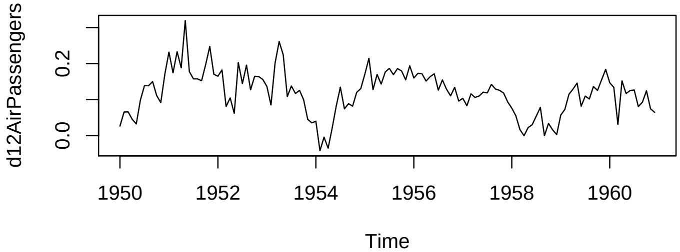


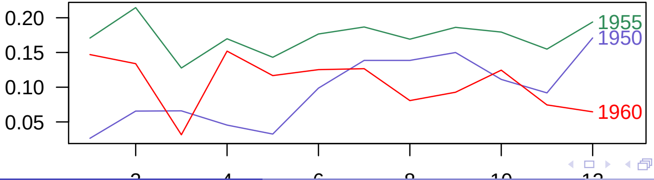


airline series: Δ

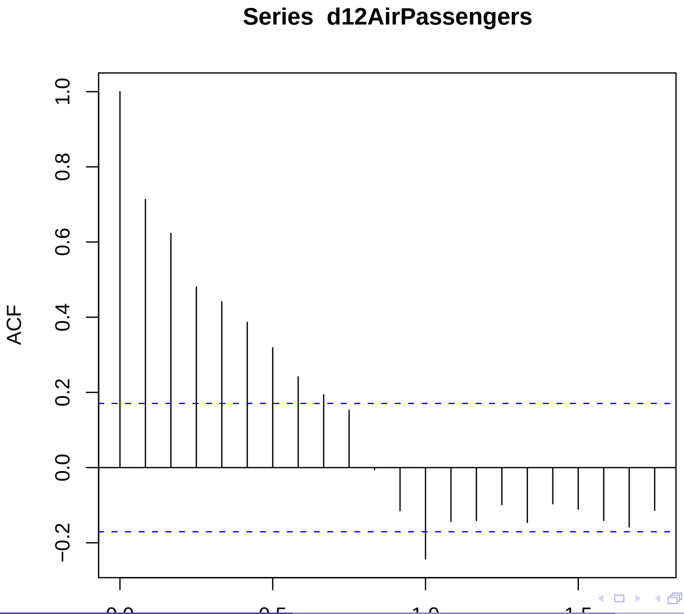


airline series: Δ_{12}

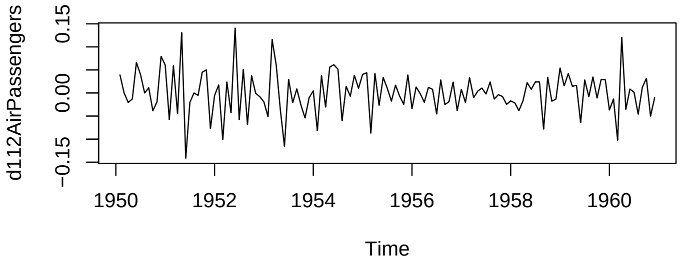


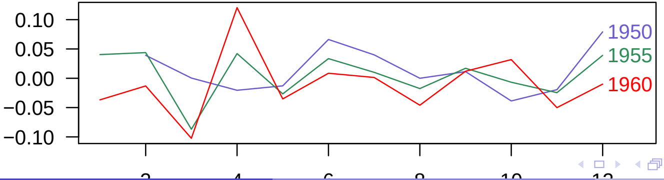


airline series: Δ_{12}

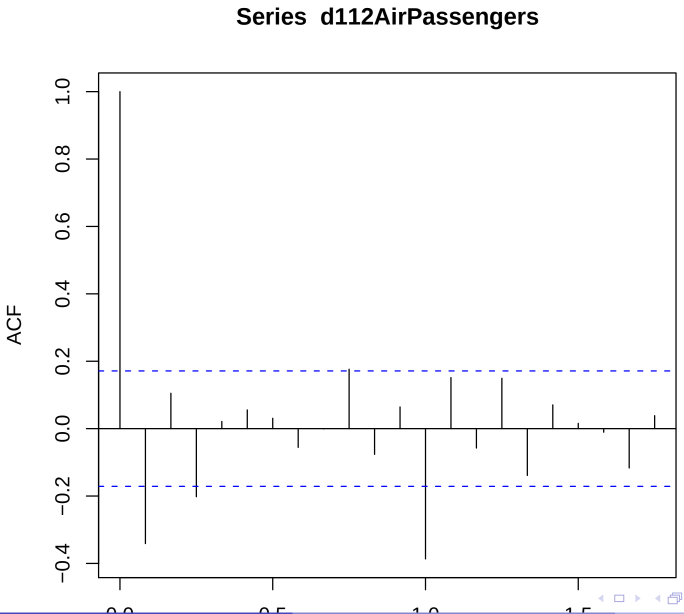


airline series: $\Delta\Delta_{12}$





airline series: $\Delta\Delta_{12}$



Box-Jenkins airline model

Box Jenkins use the sARIMA$(0,1,1)(0,1,1)_{12}$ model:

$$\Delta\Delta_{12}y_t = (1-\theta L)(1-\Theta L^{12})\varepsilon_t$$

Decomposing it gives:

$$y_t - y_{t-1} - y_{t-12} - y_{t-13} = \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-12} - \theta\Theta\varepsilon_{t-13}$$

Note that after both differencing the series is:

- Stationary
- have autocorrelation only at lags 1,11,12 and 13

Airline model

```
> seas <- (list(order = c(0, 1, 1), period = 12))
> BJ <- arima(log(AirPassengers), order = c(0, 1, 1), seasonal = seas)
> resBJ <- residuals(BJ)
> Box.test(resBJ, lag = 1, type = "Ljung")

        Box-Ljung test

data:  resBJ
X-squared = 0.0307, df = 1, p-value = 0.861

> Box.test(resBJ, lag = 12, type = "Ljung")

        Box-Ljung test

data:  resBJ
X-squared = 9.2333, df = 12, p-value = 0.6829
```

Outline

Frequency analysis

We take an empirical approach here!

Definition (Sample periodogram)

$\hat{s}_Y(\omega) = \frac{1}{2\pi} \sum_{i=0}^{T-1} \gamma_j [\cos(\omega j) - i \sin(\omega j)]$

Proposition (Equivalence 1)

$\hat{s}_Y(\omega) = \frac{1}{2\pi} \left(\gamma_0 + 2 \sum_{i=1}^{T-1} \gamma_j \cos(\omega j) \right)$

- What does it show?
- Which ω (the frequencies) do we take?

names

It seems that the periodogram is also called *spectrum* or *spectral density*.

Frequencies of the sample periodogram

Take the frequencies as:

$$\omega_j = \frac{2\pi j}{T} \qquad j = i, \ldots, [T/2]$$

Definition (period)

The period coresponding to the frequency *omega* is given by: $\frac{2\pi}{\omega}$

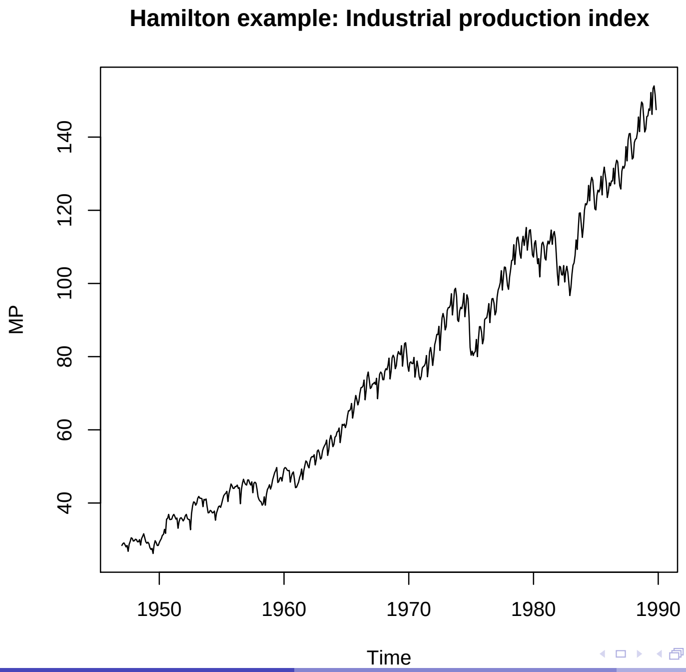
The period can be seen as the numbers of times units (months, quarters) needed to accomplish a cycle.
Low frequencies have a big period and are seen as rather "trend" elements.

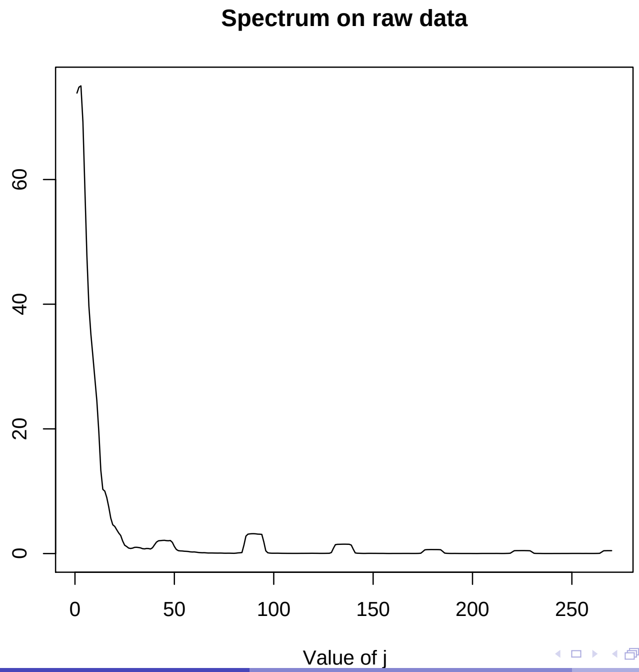
Decomposition of the variance

Proposition

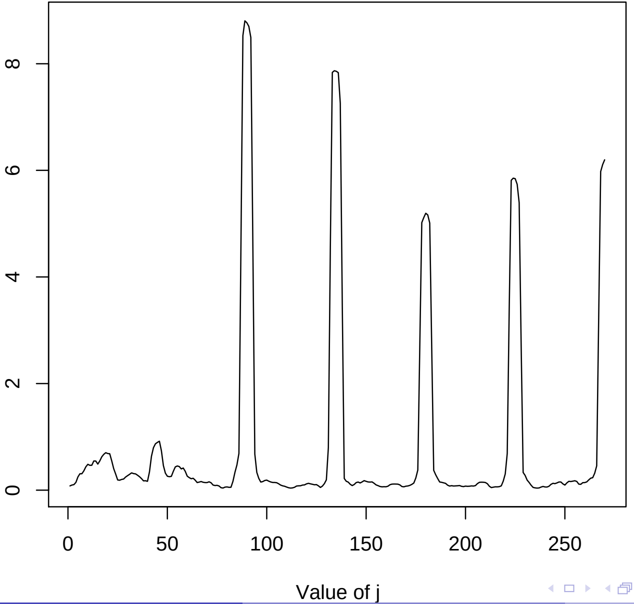
$\hat{\gamma}_0 = \int_{-\pi}^{\pi} \hat{s}_Y(\omega) d\omega = 2 \int_0^{\pi} \hat{s}_Y(\omega) d\omega$

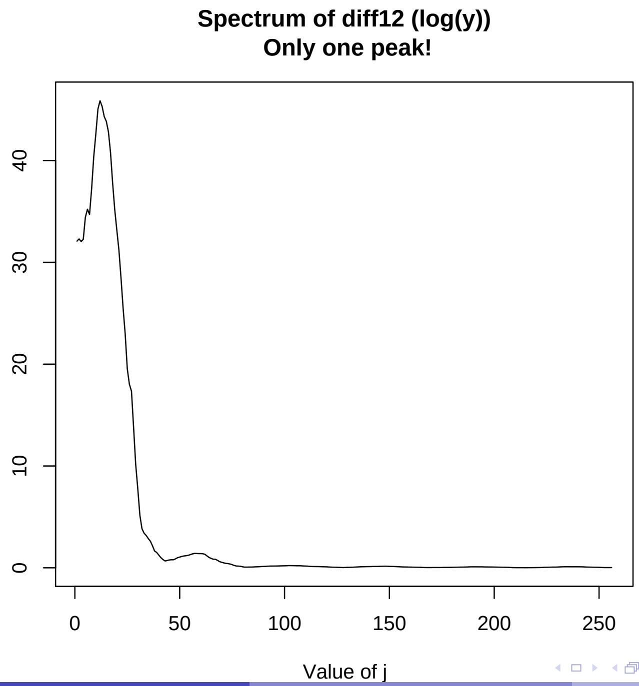
So $\int_{-\omega_j}^{\omega_j} \hat{s}_Y(\omega) d\omega$ represents the *portion of the variance of Y that could be attributed to periodic random components with frequency less than or equal to ω_j.*


Hamilton example: Industrial production index
MP
40
60
80
100
120
140
1950
1960
1970
1980
1990
Time


Spectrum on raw data
0
20
40
60
0
50
100
150
200
250
Value of j

Spectrum of diff (log(y))
Peaks are at 18 (28.5), 44 (11.7), and months 6,4,3,2.5




Spectrum of diff12 (log(y))
Only one peak!
0
10
20
30
40
0
50
100
150
200
250
Value of j

Seasonal unit roots

The polynomial $(1 - L^4)$ can be decomposed as:

$$(1 - L)(1 + L)(1 + L^2) = (1 - L)(1 + L)(1 - iL)(1 + iL)$$

It has the four roots 1,-1,i,-i and we can find their corresponding frequency!

Proposition

The frequency ω corresponding to a root is the argument of the root.

Definition

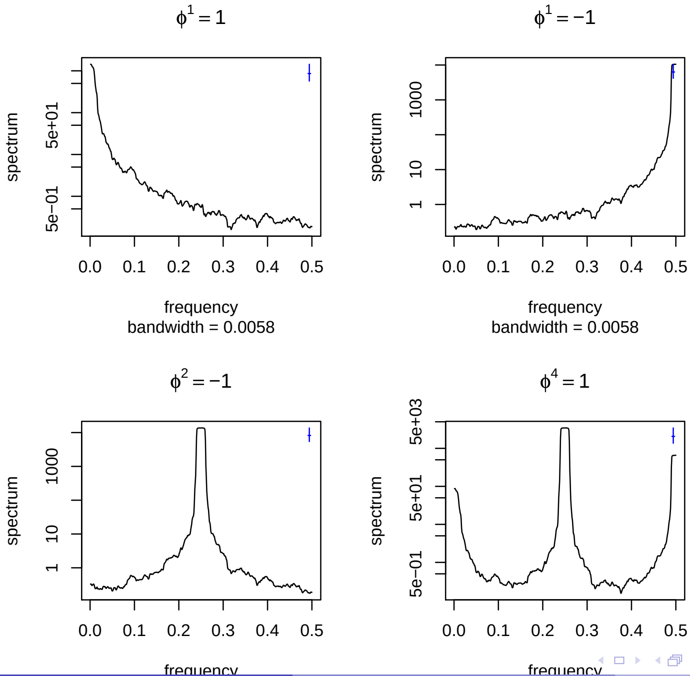
The argument of a complex number is the angle in the polar representation and is given by: $\omega = \cos^{-1}(a/R) = sin^{-1}(b/R)$, a is the real part, b the imaginary and R the modulus.

Seasonal unit root

Proposition

The number of cycles per year is given by $\frac{S\omega}{2\pi}$

root	f	period	Annual occurence
1	0	no	no
-1	π	2 quarters	2
i,-i	$\pi/2, -\pi/2$	4 quarters	1


$\phi^1 = 1$
$\phi^1 = -1$
$\phi^2 = -1$
$\phi^4 = 1$
spectrum
5e−01
5e+01
5e+03
1
10
1000
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0058

The HEGY test

The HEGY (1990) test decompose the polynomial complexically and runs an auxilliary regression:

$$y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

As example y_{1t} is $(1 + L + L^2 + L^3)x_t$.
Thus one has the following hypotheses:

- Nonseasonal root (1): $\pi_1 = 0$
- Seasonal bi-annual root (-1): $\pi_2 = 0$
- Seasonal annual root (i,-i): $\pi_3 = 0 \cap \pi_4 = 0$

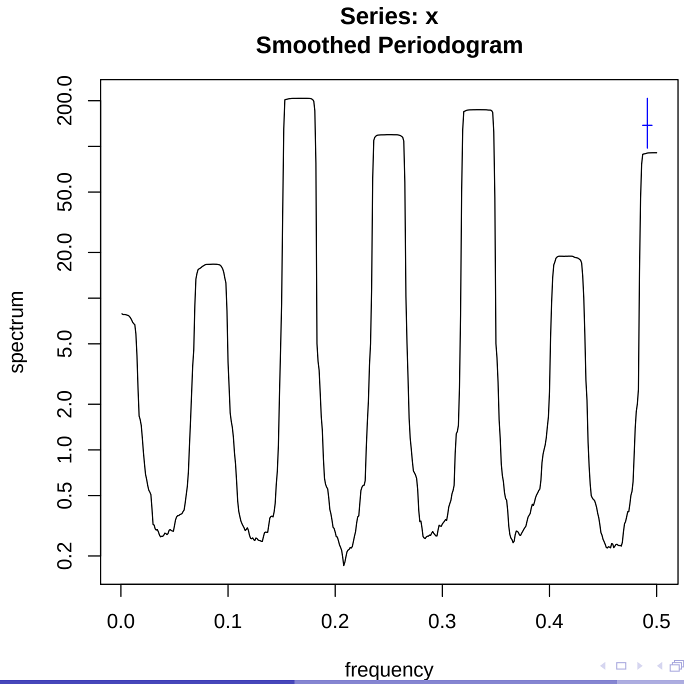
The tabulated distribution depends on whether there is intercept/trend/seasonal dummy.

HEGY for monthly series

The HEGY test has been extended for monthly series (12 roots!) by Franses (1990) and Beaulieu and Miron (1993)

The roots are the same as HEGY (1,-1, i,-1) plus $\pm 1/2(1 \pm \sqrt{3}i)$, $\pm 1/2(\sqrt{3} \pm i)$

root	freq	cycles per year
1	π	6
i	$\pi/2$	3
i	$-\pi/2$	9
$-1/2(1 + \sqrt{3}i)$	$-2\pi/3$	8
$-1/2(1 - \sqrt{3}i)$	$2\pi/3$	4
$1/2(1 + \sqrt{3}i)$	$\pi/3$	2
$1/2(1 - \sqrt{3}i)$	$-\pi/3$	10
$-1/2(\sqrt{3} + i)$	$-5\pi/6$	7
$-1/2(\sqrt{3} - i)$	$5\pi/6$	5
$1/2(\sqrt{3} + i)$	$\pi/6$	1
$1/2(\sqrt{3} - i)$	$-\pi/6$	11


Series: x
Smoothed Periodogram
spectrum
0.2
0.5
1.0
2.0
5.0
20.0
50.0
200.0
0.0
0.1
0.2
0.3
0.4
0.5
frequency

HEGY tests with R

The HEGY test and its extension to monthly data ara available in R in:

```
> library(uroot)
> data(AirPassengers)
> lairp <- log(AirPassengers)
> test <- HEGY.test(wts = lairp, itsd = c(1, 1, c(1:11)), regvar = 0,
+     selectlags = list(mode = "bic", Pmax = 12))
> test@stats
```

```
                Stat.   p-value
tpi_1       -2.577797 0.1000000
tpi_2       -4.396433 0.0100000
Fpi_3:4     18.519107 0.1000000
Fpi_5:6      4.823309 0.0100000
Fpi_7:8      8.656624 0.1000000
Fpi_9:10     7.119685 0.0494419
Fpi_11:12    2.854972 0.0100000
Fpi_2:12    18.373828        NA
Fpi_1:12    19.336146        NA
```

monthly HEGY test

> *The mechanical application of the seasonal difference filter is likely to produce serious misspecification in many instances. The evidence presented here indicates that unit roots are often absent at some or all of the seasonal frequencies, so empirical researchers should check for their presence (using procedures such as the one discussed above) rather than imposing them at all seasonal frequencies a priori.*

Beaulieu, Miron (1993)

Stationarity as a null

Canova and Hansen take opposite approach: H_0 is stationarity, with deterministic seasonality.
From:

$$y_t = \alpha y_{t-1} + \sum_{i=1}^{S-1} D_{it}\beta_i + \varepsilon_t$$

The idea is (provided stationarity, i.e. $|\alpha| < 1$) to test for instablity of the β_i parameters (see lecture 5) as the KPSS test does (lecture 4):

$$y_t = \alpha y_{t-1} + \sum_{i=1}^{S-1} D_{it}\beta_{it} + \varepsilon_t$$

$$\beta_{it} = \beta_{it-1} + u_t$$

and test if $\text{Var}(u_t) = 0$
The test can also be applied to only a subset of dummies.

```
> CH.test(wts = AirPassengers, frec = c(1, 1, 1, 1, 1, 1), f0 = 1,
+     DetTr = FALSE)

  ------ - ------ ----
  Canova & Hansen test
  ------ - ------ ----

  Null hypothesis: Stationarity.
  Alternative hypothesis: Unit root.
  Frequency of the tested cycles: pi/6 , pi/3 , pi/2 , 2pi/3 , 5pi/6

  L-statistic: 1.836
  Lag truncation parameter: 13

  Critical values:

 0.10 0.05 0.025 0.01
 2.49 2.75  2.99 3.27
```

Outline

Outline

type of seasonality

Additive:

$$X_t = C_t + S_t + TD_t + H_t + I_t$$

Multiplicative:

$$X_t = C_t \times S_t \times TD_t \times H_t \times I_t$$

Where:

- C is the trend-cycle component
- S is the seasonal component
- TD is trading-day effect
- H is the holiday effect
- I is the irregular component

the specs

- series{}
- arima{}: ARIMA given by user
- automdl{}: automatic choice of the ARIMA
- check{}: print tests on the regARIMA
- estimate {} print tests on the regARIMA
- x11{}: the seasonal adjustment procedure

X-12-ARIMA

The regARIMA option

$$\Phi(L^s)\phi(L)\Delta^d\Delta_s^D(y_t - \sum x_{it}) = \Theta(L^s)\theta(L)\varepsilon_t$$

x_i can be:

- Constant/trend pattern:
  - Constant if $D = d = 0$ (then equals mean of the series)
  - Sort of trend if $D, d \neq 0$
- Fixed seasonal pattern:
  - seasonal dummies
  - Sin-cos functions
- Level shift (structural break)
- Temporary change
- External variable

Deterministic vs stochastic seasonality

If arg season is given, you can't use seasonal differencing (i.e. D=0)

X-12-ARIMA: ARIMA fitting

> *Estimation and identification can be carried out by following well-established procedures that rely on examination of ACF and PACF of y_t and its differences.*

If regARIMA provided: ACF and PACF are different, so other methodology.

X-12-ARIMA: ARIMA estimation

Estimator provided: exact MLE.

Inference: it seems that no inference is made on ARIMA and regARIMA for the variables added by the user, but can be extracted (see page 42), you must use the spec check{}

Otherwise, one can use usual information criteria to select the model AIC, AAIC, BIC, Hannan-Quin

p. 50: aictest argument of the regression and x11regression specs is used to automatically decide for or against the inclusions of certain regressors (see Sections 7.13 and 7.18 of the X-12-ARIMA Reference Manual).

Spec check{}

The check spec is used to produce various diagnostic statistics using the residuals from the fitted model. To check for autocorrelation, X-12-ARIMA can produce:

- ACFs and PACFs of the residuals (with standard errors)
- Ljung and Box (1978) summary Q-statistics
- basic descriptive statistics of the residuals
- histogram of the standardized residuals

The spec autmomdl{}

automdlprint = (none bestfivemdl autochoice) savelog = automodel
Model selection procedure is adapted from TRAMO

1. **default model estimation**: a default model is estimated, initial outlier identification and regressor tests are performed, and residual diagnostics are generated;
2. **identification of differencing orders**: empirical unit root tests are performed to determine the orders
3. **identification of ARMA model orders**: an iterative procedure is applied
4. **comparison of identified model with default model**: the identified model is compared to the default model
5. **final model checks**: where the final model is checked for adequacy

automdl: step 1

Test from the airline model:

- trading day, Easter (AIC?)
- user-defined regressors (AIC?)
- const in regARIMA (t-test)
- Ljung-Box test of the residuals

Model selected should perform better than this one.
Manual 3, page 70

automdl: step 2 and 3

Step 2: identification of d,D Use empirical unit root tests: estimate $(200)(100)_s$ and see if the roots of the AR polynomial are outside the unit circle.
Step 3: identification of p,q,P,Q Multi-step BIC procedure
If model prefered is different than airline model, then make all tests (easter day, user-defined var, LB stat)

automdl: step 4

Compares with the airline model:

- Number of outliers is less
- Ljung-Box statistic is better
- Some empirical rules

  The program then tests to see if the preferred model is acceptable. The confidence coefficient of the Ljung- Box Q statistic is used as the criterion.

automdl: step 5 final check

Checks:

- If sum of AR is outside unit root circle
- Is unit root in non-seasonal MA polynomial (still stationary but not invertible)
- If constant in model is significant (when not previously given)
- If ARMA coefficients are significants (individually)

R package forecast

Function auto.arima

- Canova and Hansen (1995) test for D
- Test as stationarity as null hypothesis (KPSS) for d

R implementation

To run this Rnw file you will need:

- Package uroot
- Data file of Hamilton: Ham.txt (same folder as .Rnw file)
- (Optional) File Sweave.sty which change output style: result is in blue, R commands are smaller. Also in same folder as .Rnw file.