Time series analysis using ARFIMA

Frank Ebert
Ebert Beratung und Innovationen GmbH, Wedel

German Stata User Group Meeting – 26. June 2015 – Nürnberg

Time series analysis using ARFIMA

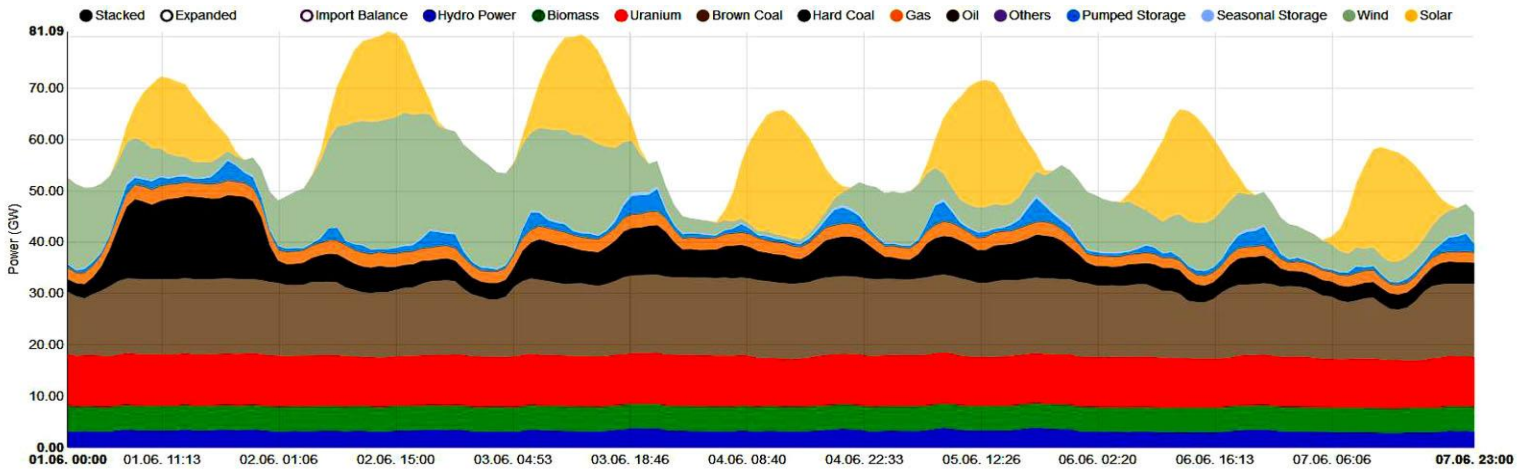
Background	Time series analysis
Stata ARFIMA	Results

Background – production of electric energy in Germany

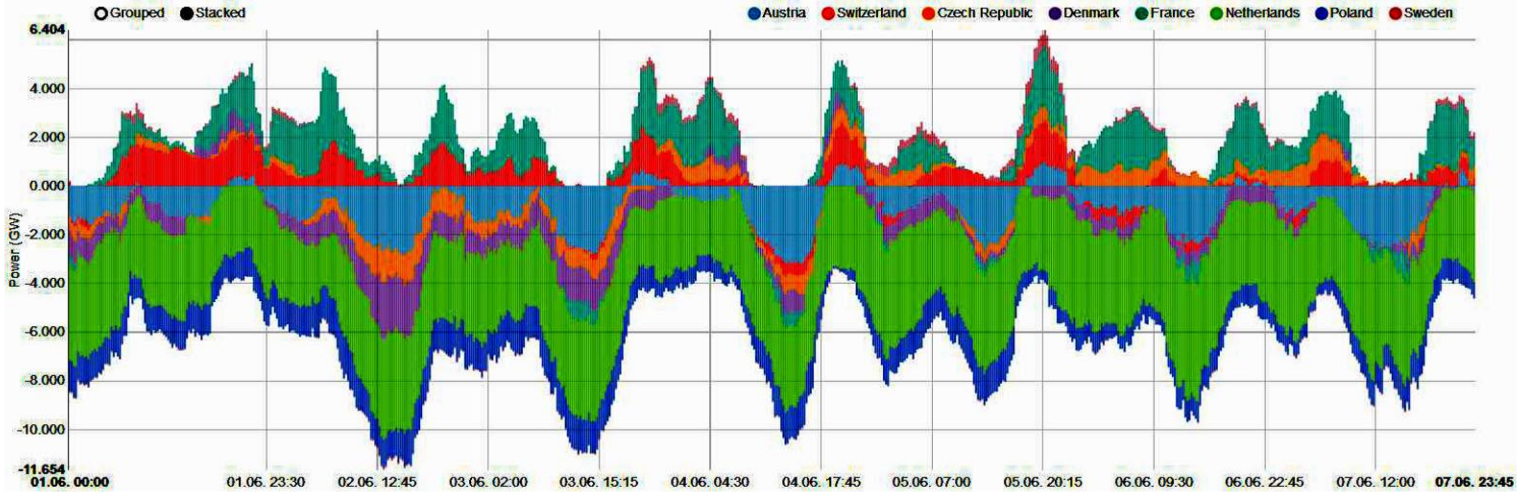
Production of electric energy
Germany

Week 25 / 2015

(Average)



Import / Export



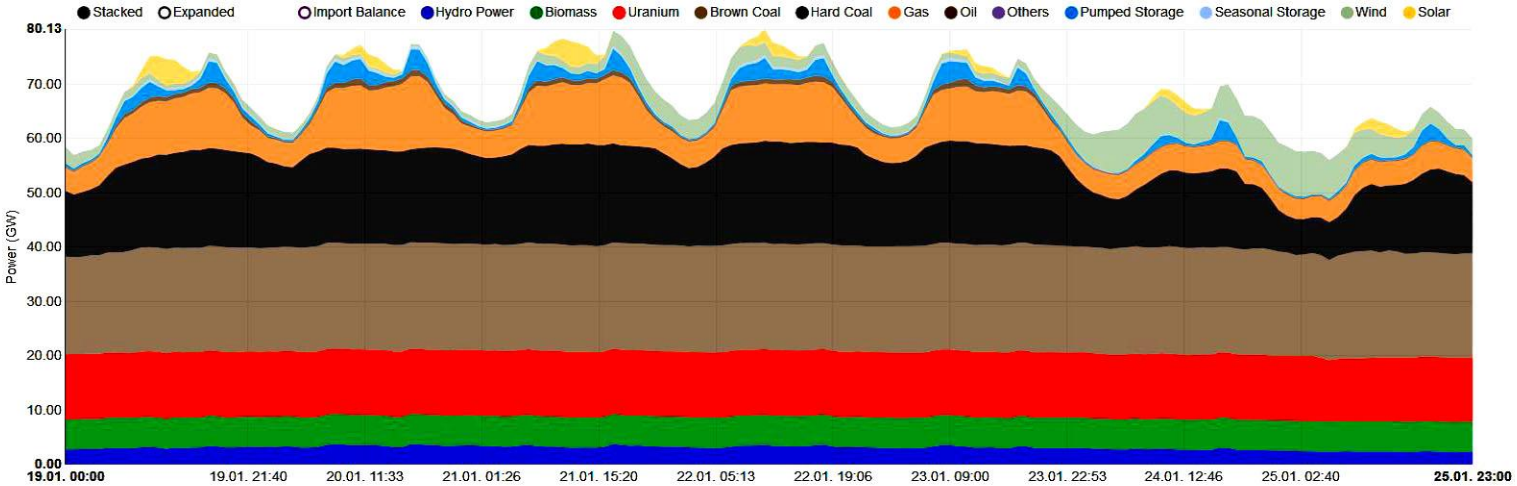
Source: Fraunhofer-Institut für Solare Energiesysteme ISE, https://www.energy-charts.de/power.htm

Background – production of electric energy in Germany

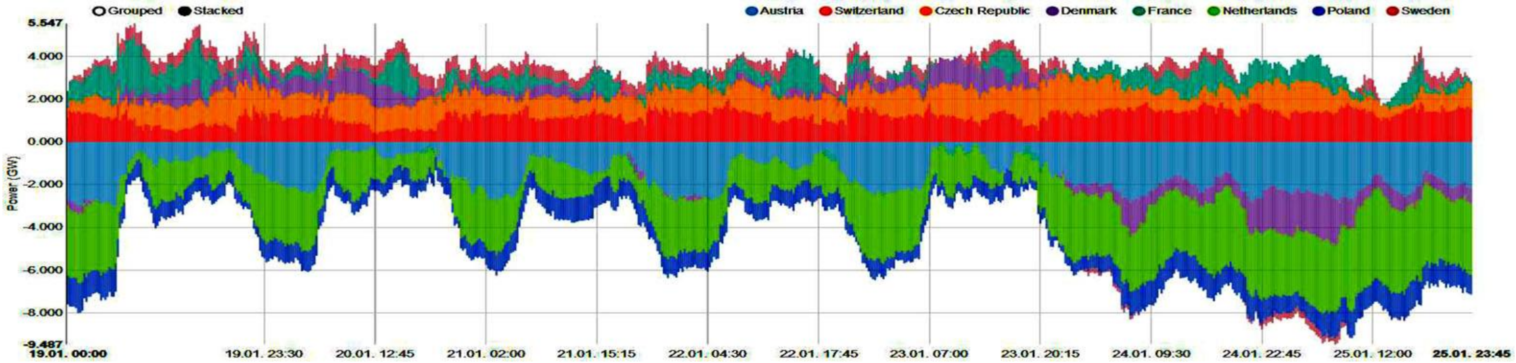
Production of electric energy
Germany

Week 04 / 2015

„Dark“



Import / Export



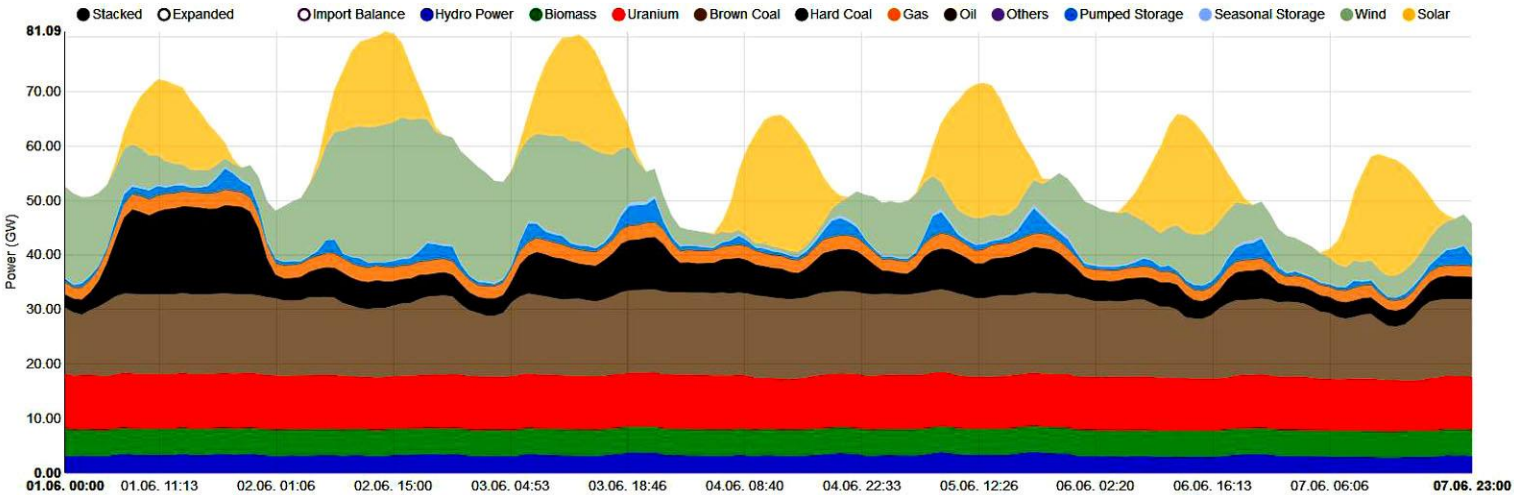
Source: Fraunhofer-Institut für Solare Energiesysteme ISE, https://www.energy-charts.de/power.htm

Background – production of electric energy in Germany

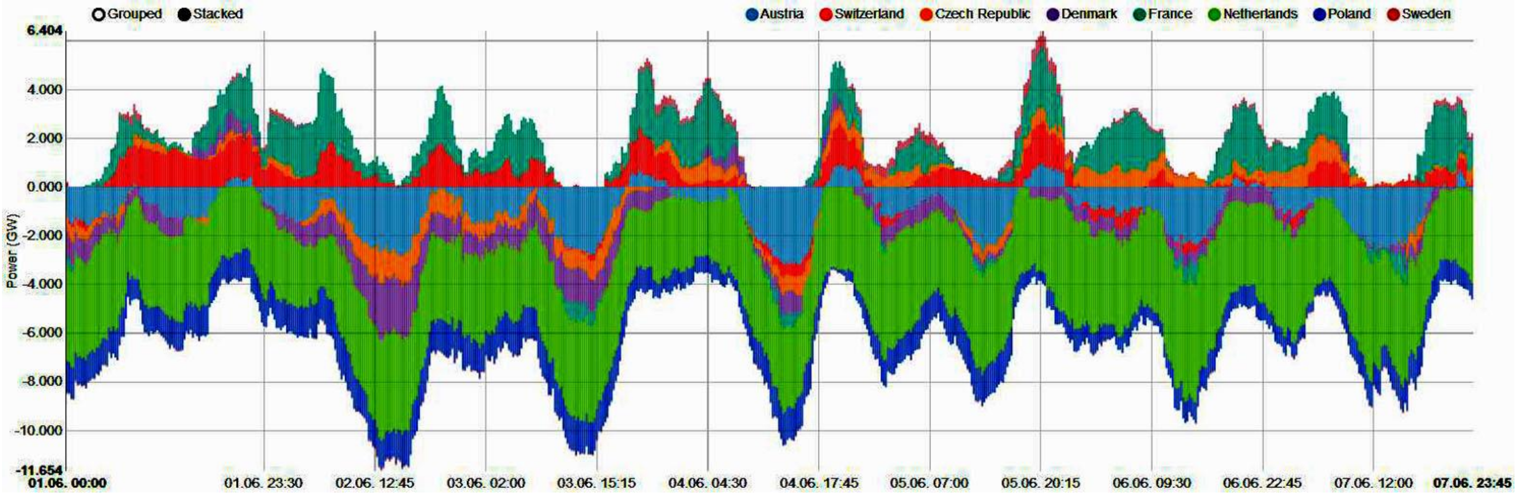
Production of electric energy
Germany

Week 23 / 2015

„Sunny“



Import / Export



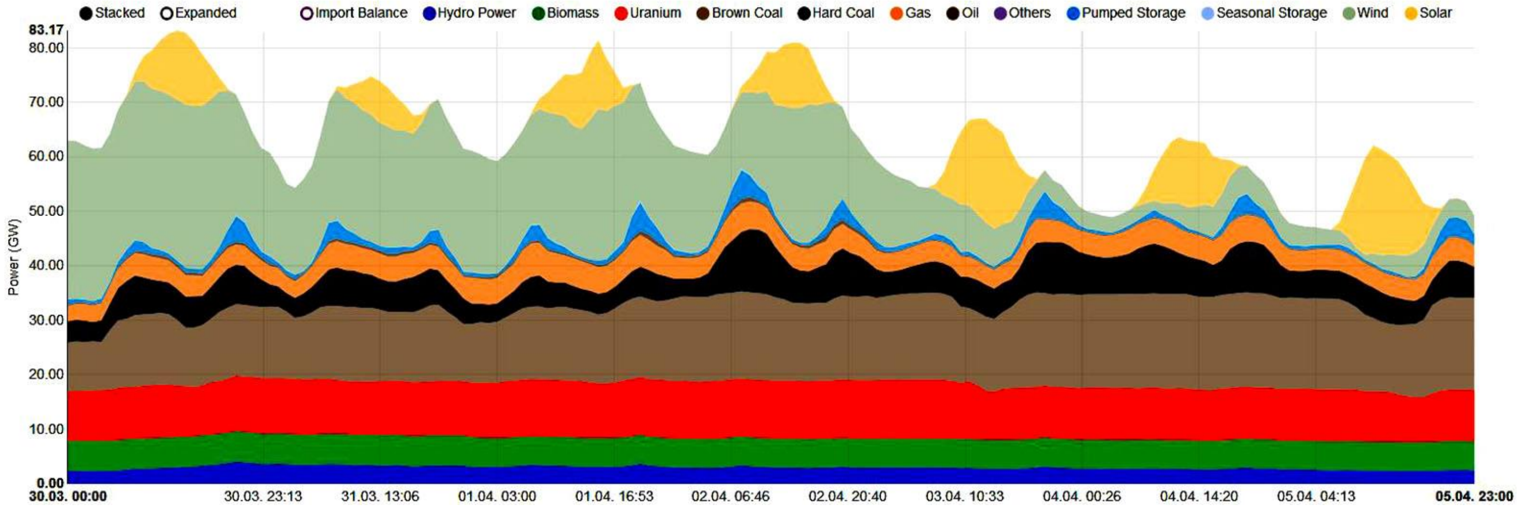
Source: Fraunhofer-Institut für Solare Energiesysteme ISE, https://www.energy-charts.de/power.htm

Background – production of electric energy in Germany

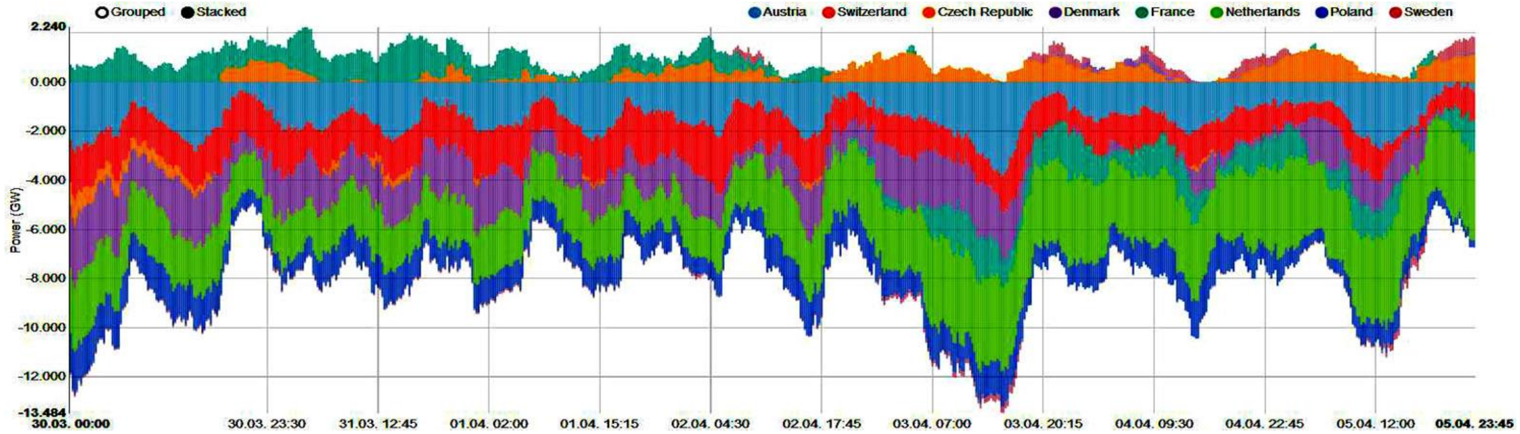
Production of electric energy Germany

Week 14 / 2015

„Windy“

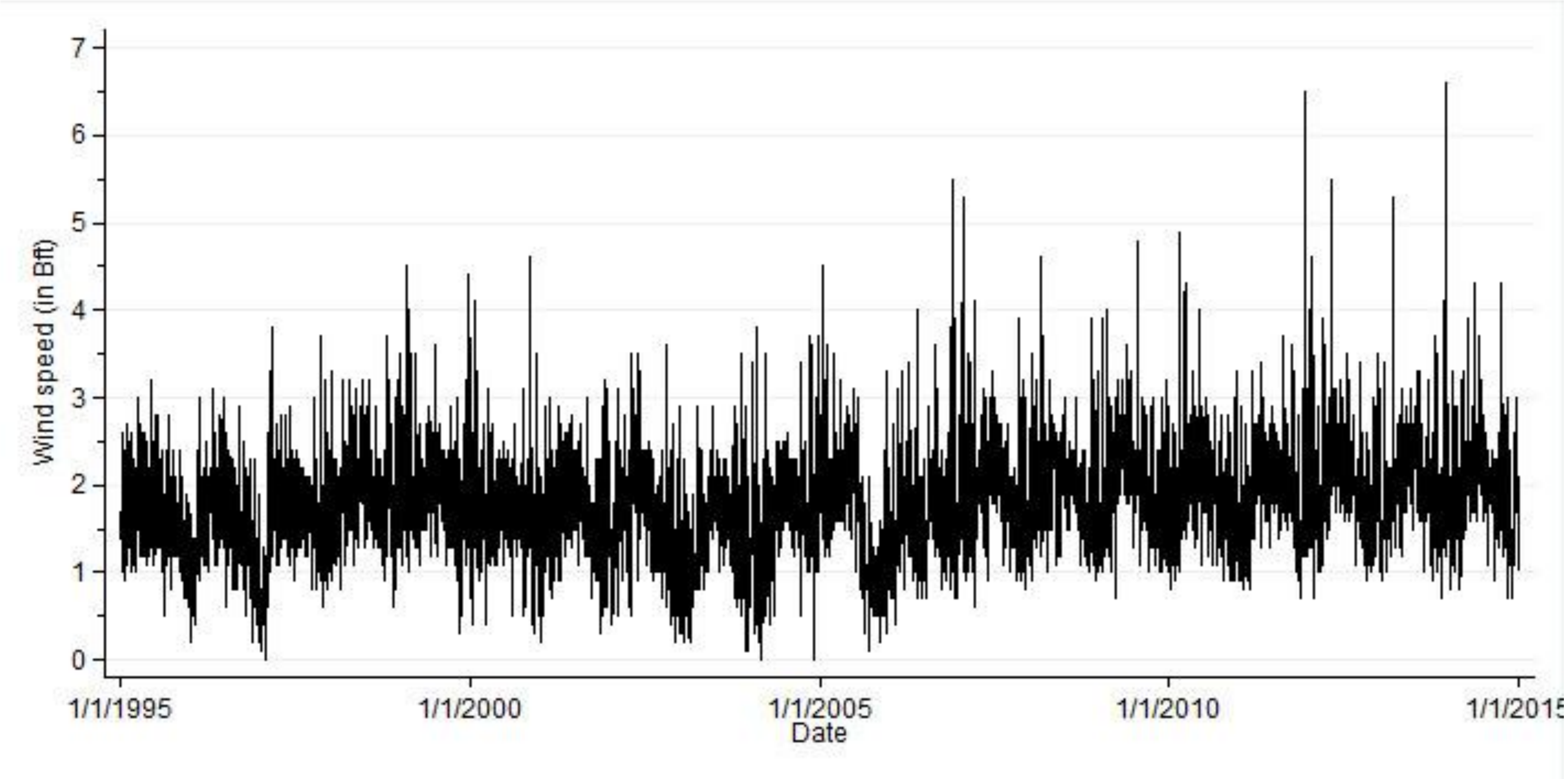


Import / Export



Source: Fraunhofer-Institut für Solare Energiesysteme ISE, https://www.energy-charts.de/power.htm

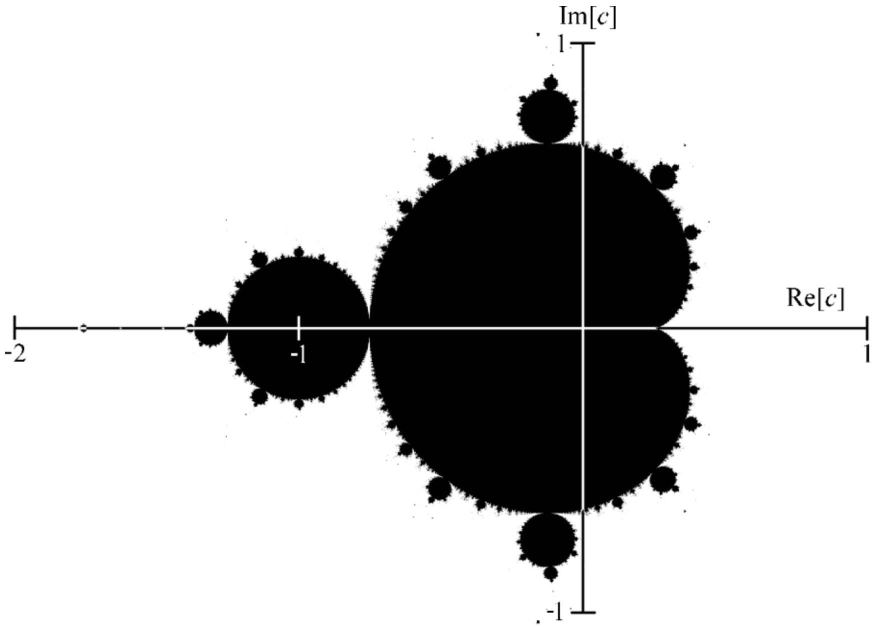
Background –
Possibly fractal character of weather time series



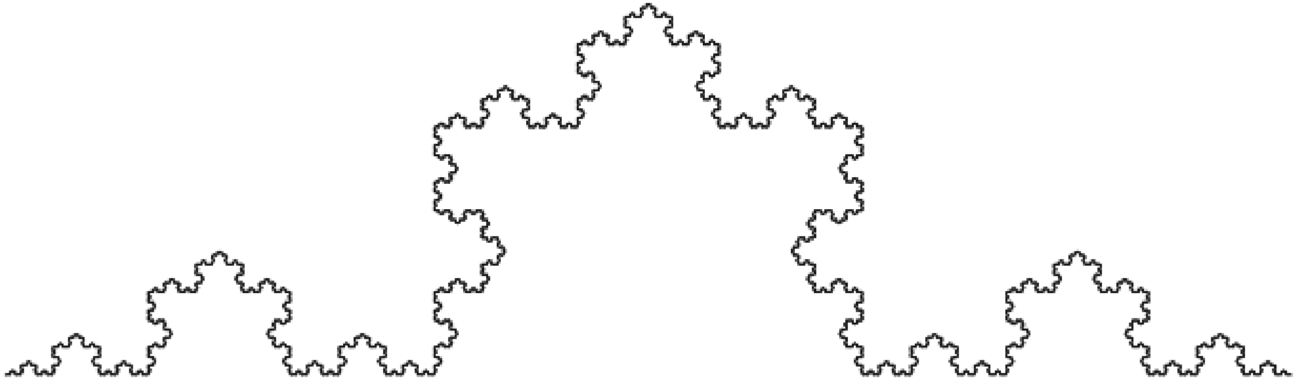
Average daily wind speed (Oberstdorf)

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Background – classic fractals



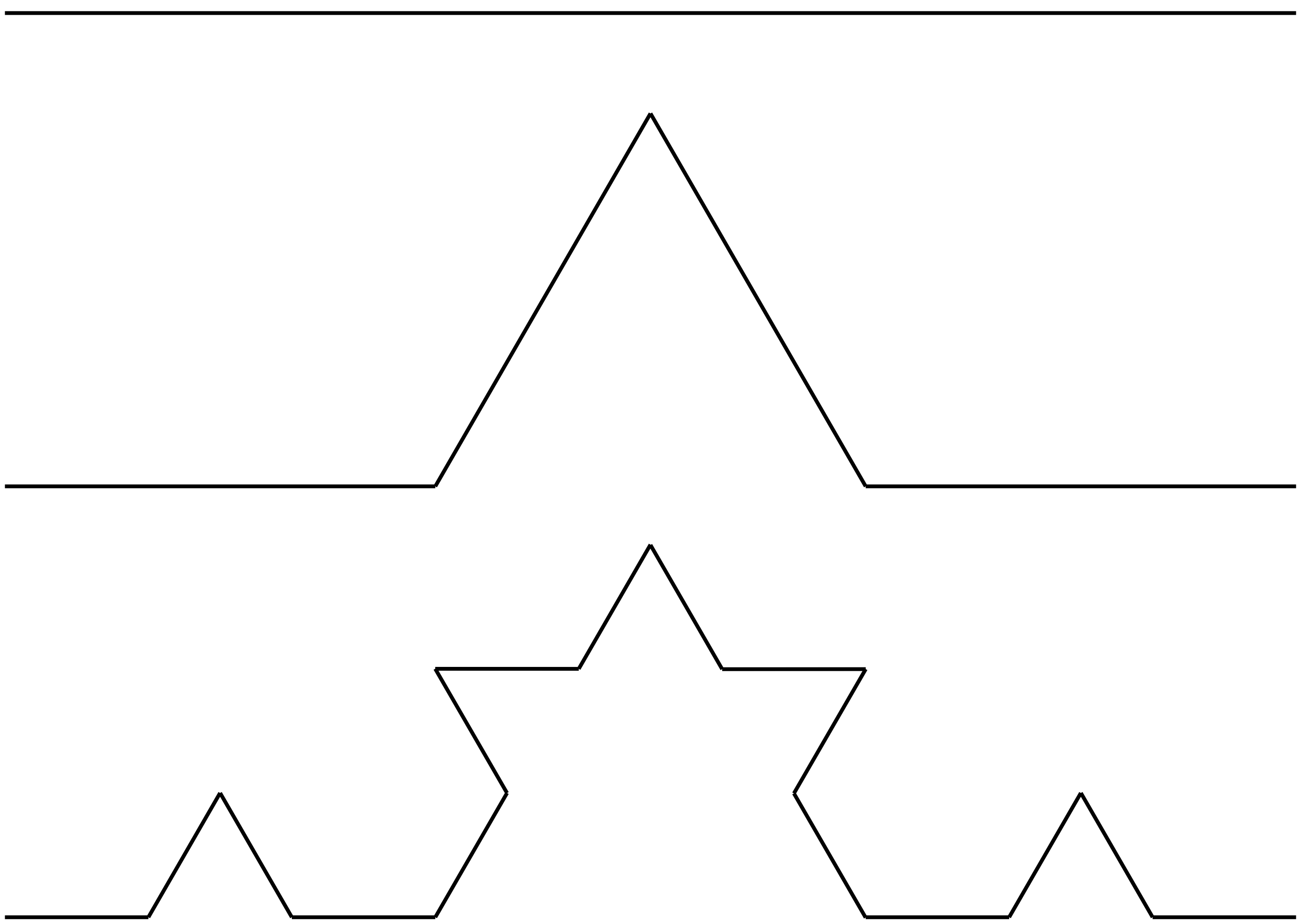
Mandelbrot set



Von Koch curve

Source: https://upload.wikimedia.org/wikipedia/commons/5/56/Mandelset_hires.png

Background – classic fractals



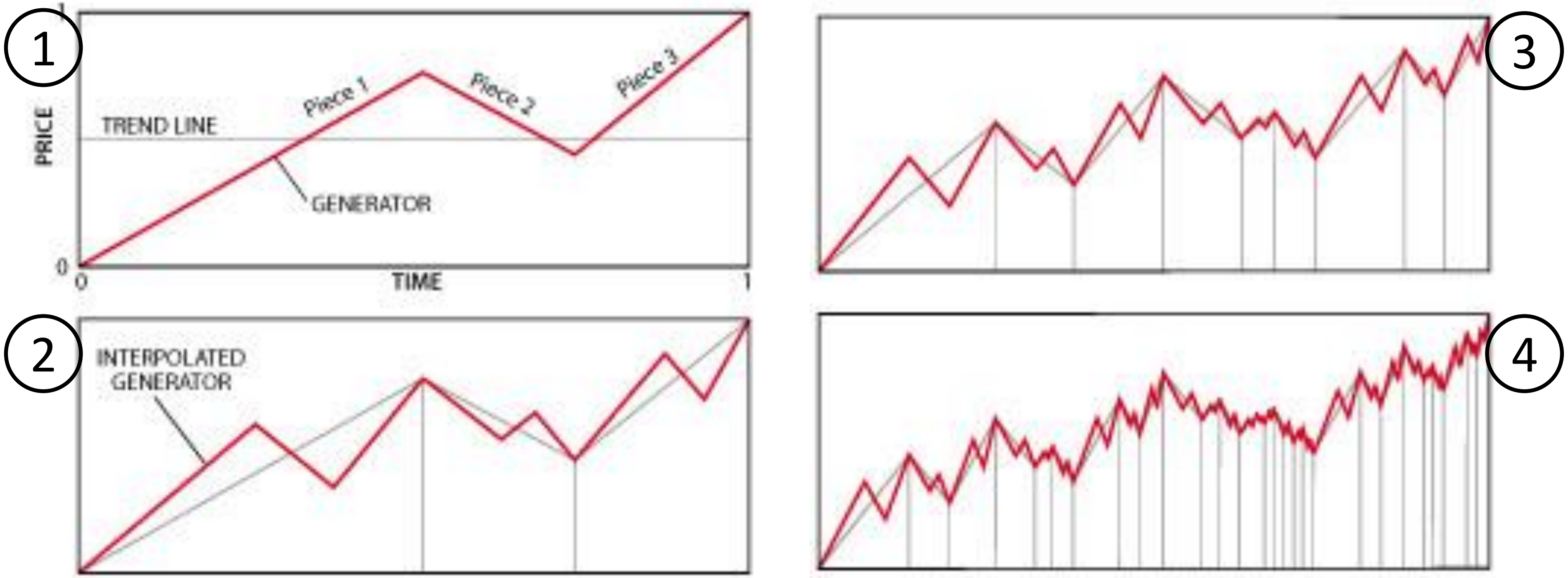
Construction of the Von Koch curve

Hausdorff-Dimension
log (4) / log (3) ≈ 1.26

Background – fractals

1 THREE-PIECE FRACTAL GENERATOR (*top*) can be interpolated repeatedly into each piece of subsequent charts (*bottom three diagrams*). The pattern that emerges increasingly resembles market price oscillations. (The interpolated generator is inverted for each descending piece.)

Construction of a non-random fractal



Source: Mandelbrot, B. B. (1999). A multifractal walk down Wall Street. *Scientific American. 1999(2)*, pp. 70-73

Time series analysis using ARFIMA

Background

Time series analysis

Stata ARFIMA

Results

Time series analysis – AR(p) process

Autoregressive process of order p $\boldsymbol{(X_t)_{t\in T}}$

$X_t \;=\; \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + u_t$

$(u_t)_{t\in T}$ white noise process

u_t independent random variables with uniform probability distribution

$T = \mathbb{N}$ or $T = \mathbb{Z}$

Backshift operator

$B^k X_t = X_{t-k}\,, \quad k = 1, 2, \ldots$

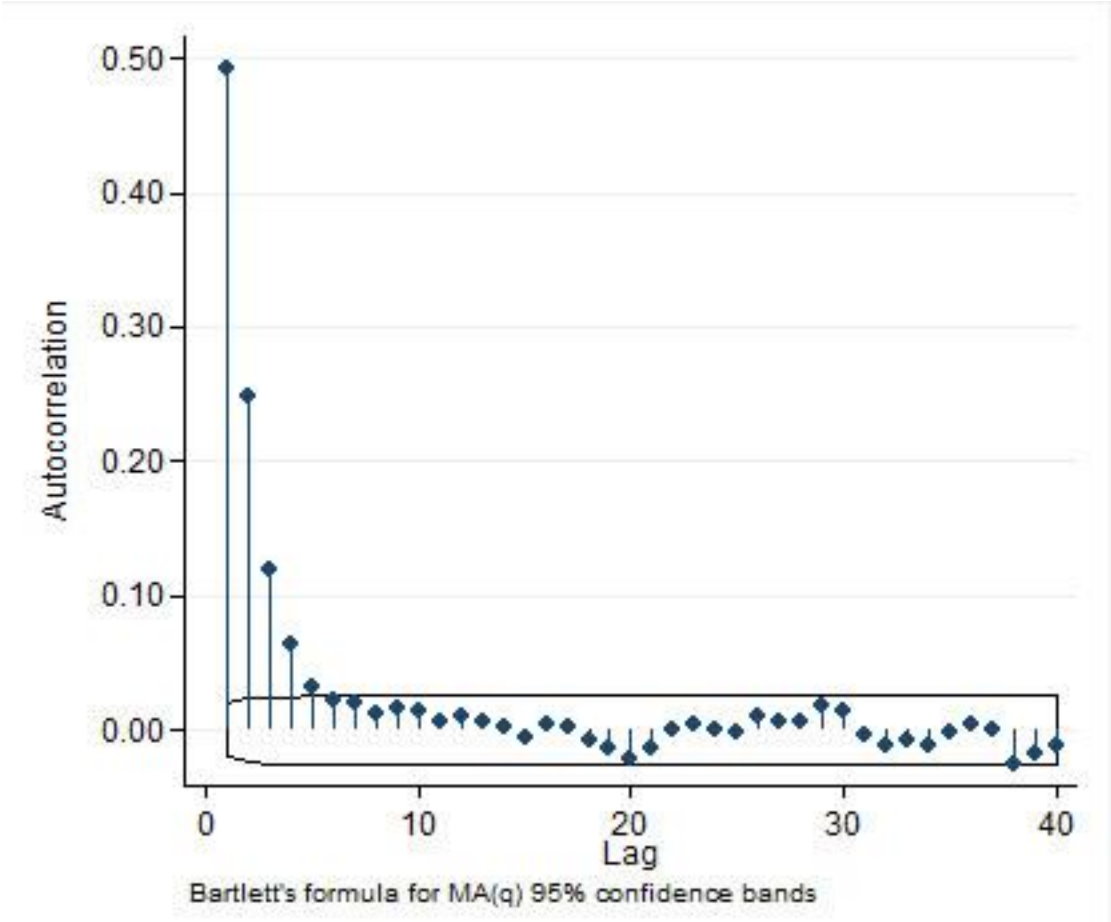
Alternative notion

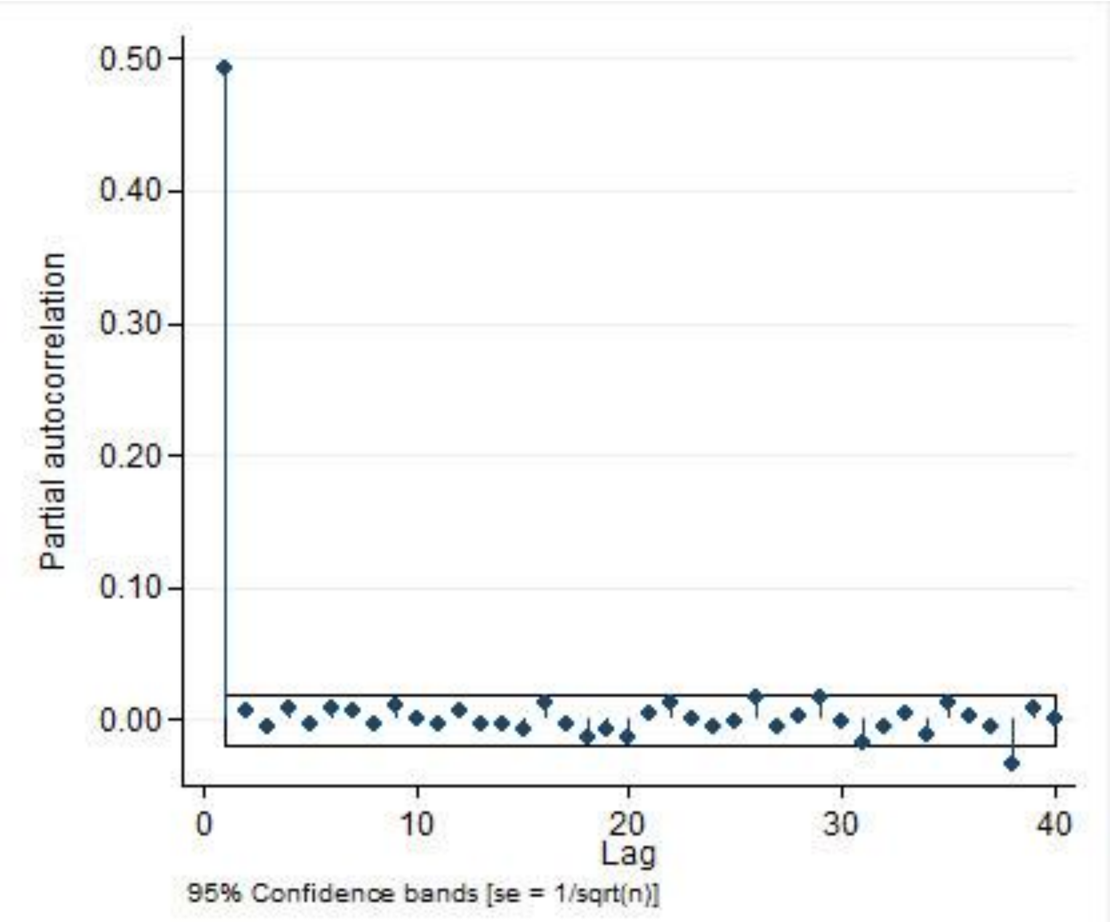
$(1 \,-\, \phi_1 B \,-\, \phi_2 B^2 \,-\cdots-\, \phi_p B^p)X_t \;=\; u_t$

Shorter

$\Phi_p(B)X_t \;=\; u_t$

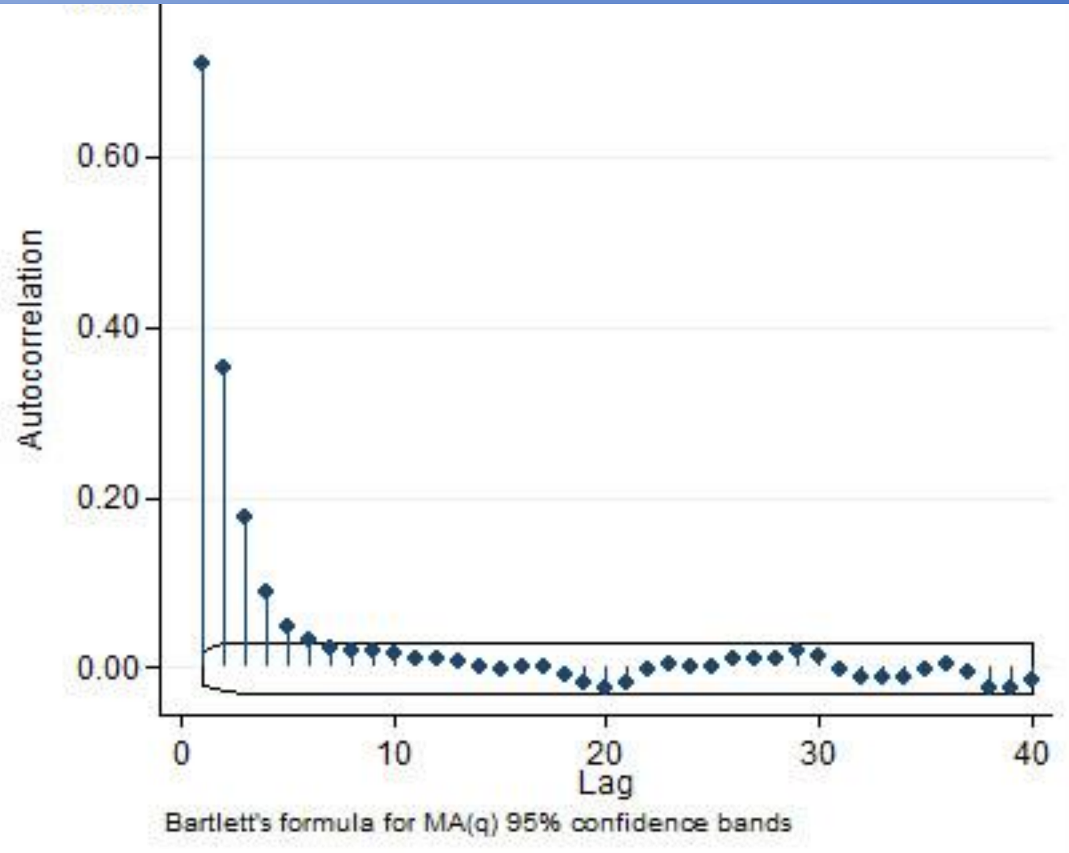
Time series analysis – AR(1) process



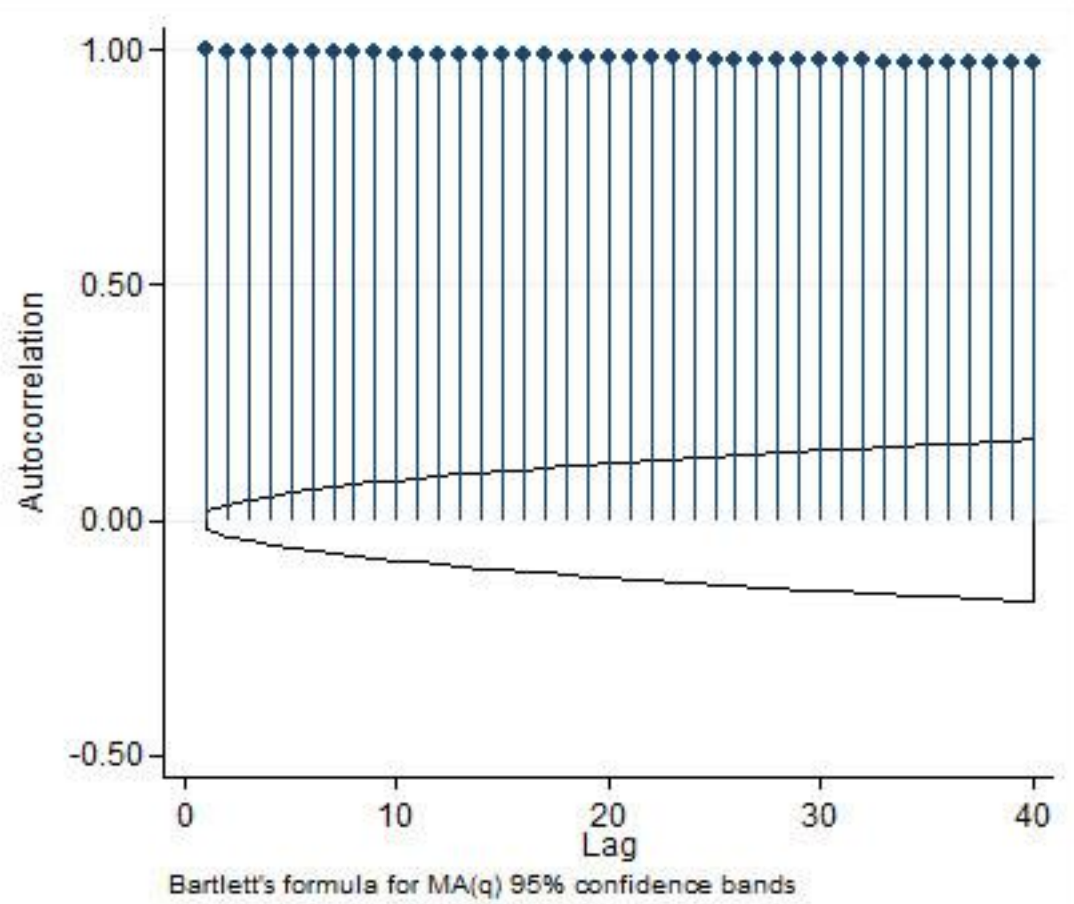


Autocorrelation function (ACF) and partial autocorrelation function (PACF) of an AR(1) process with p = 0.5 (10,000 simulated values)

Time series analysis – AR(1) process



ARMA(1, 1) process with p=q=0.5



Brownian motion (independent inreases)

Time series analysis – MA(q) process

Moving average process of order q $(\boldsymbol{X_t})_{\boldsymbol{t \in T}}$

$$X_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_p u_{t-q}$$

Time series dependent on past estimation errors

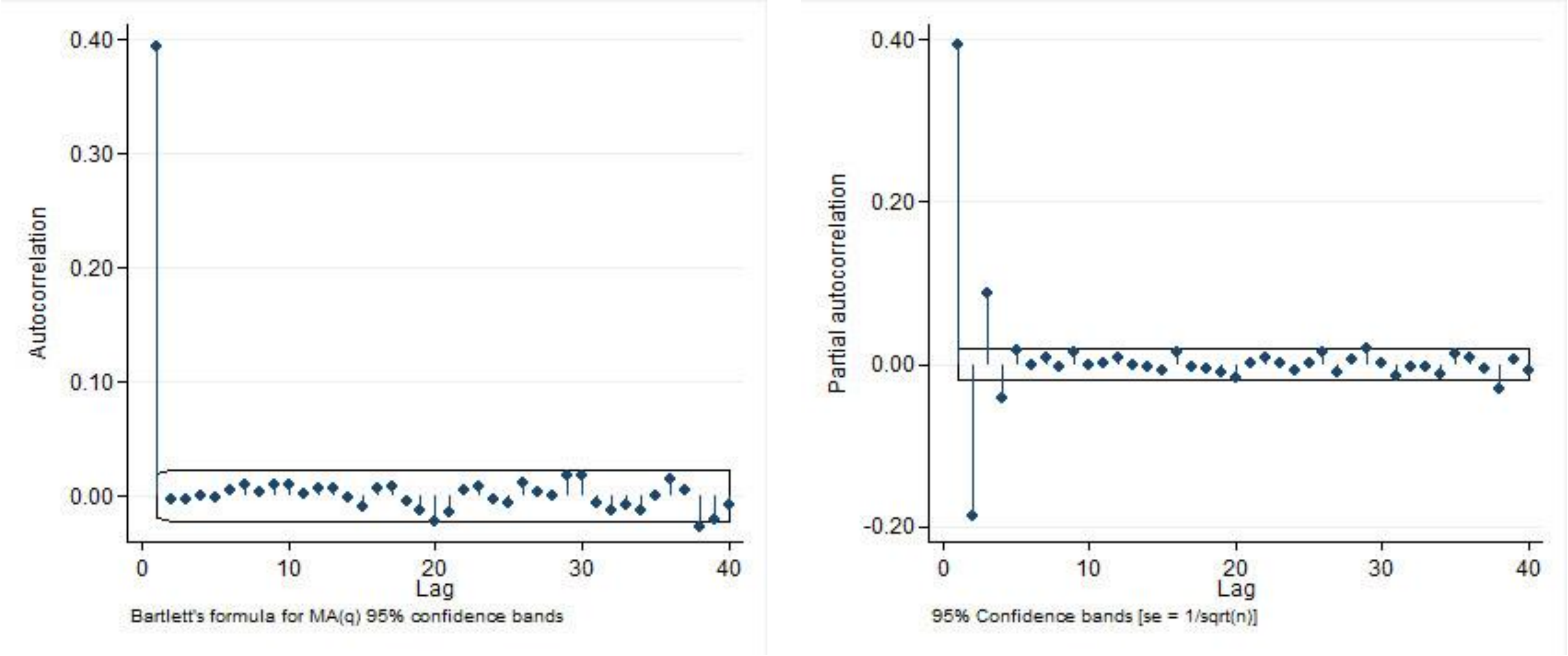
Alternative notion

$$\left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^q\right) u_t = X_t$$

Shorter

$$\Theta_q(B) u_t = X_t$$

Time series analysis – MA(1) process



ACF and PACF of a MA(1) process with q=0.5 (10,000 simulated values)

Time series analysis – ARMA(p, q) process

Autoregressive moving average process of order p, q $(\boldsymbol{X_t})_{\boldsymbol{t \in T}}$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_p u_{t-q}$$

Alternative notion

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^q) u_t$$

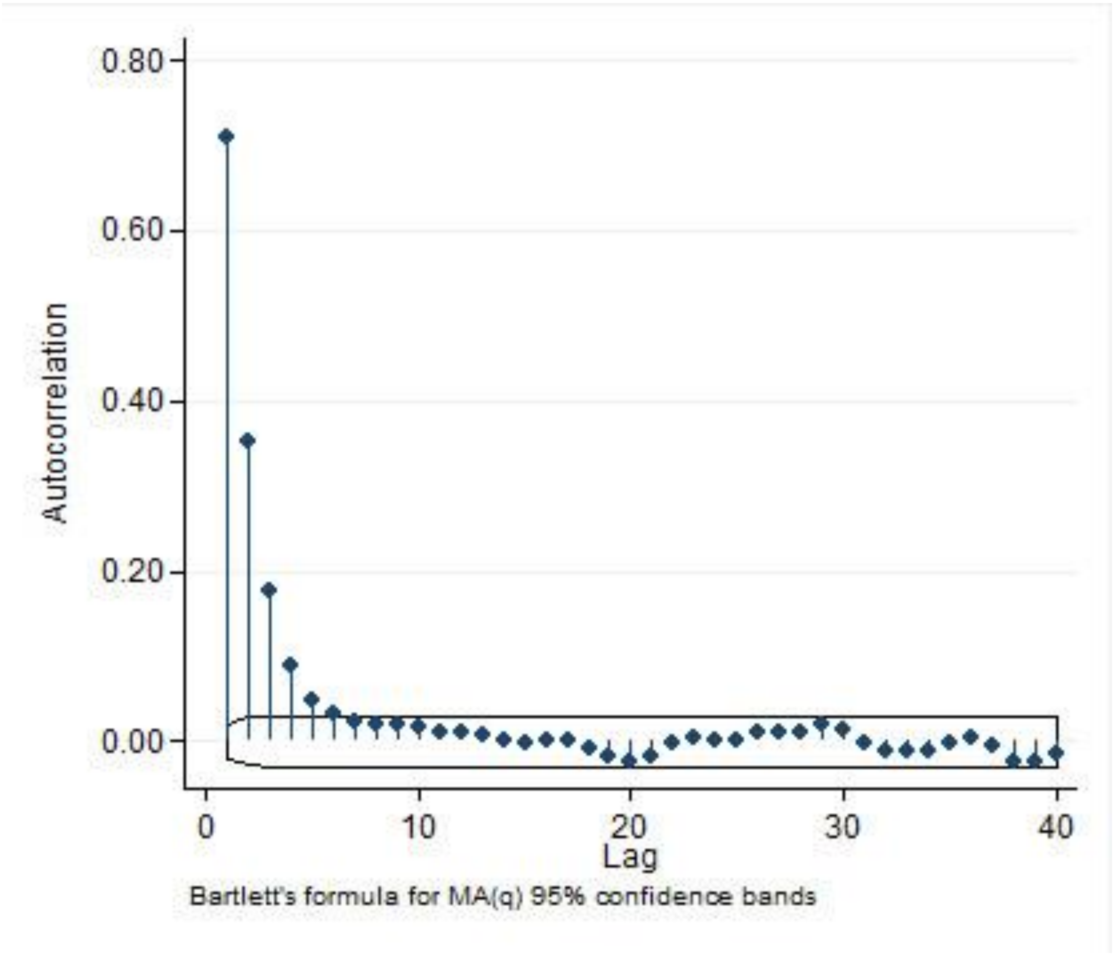
Shorter

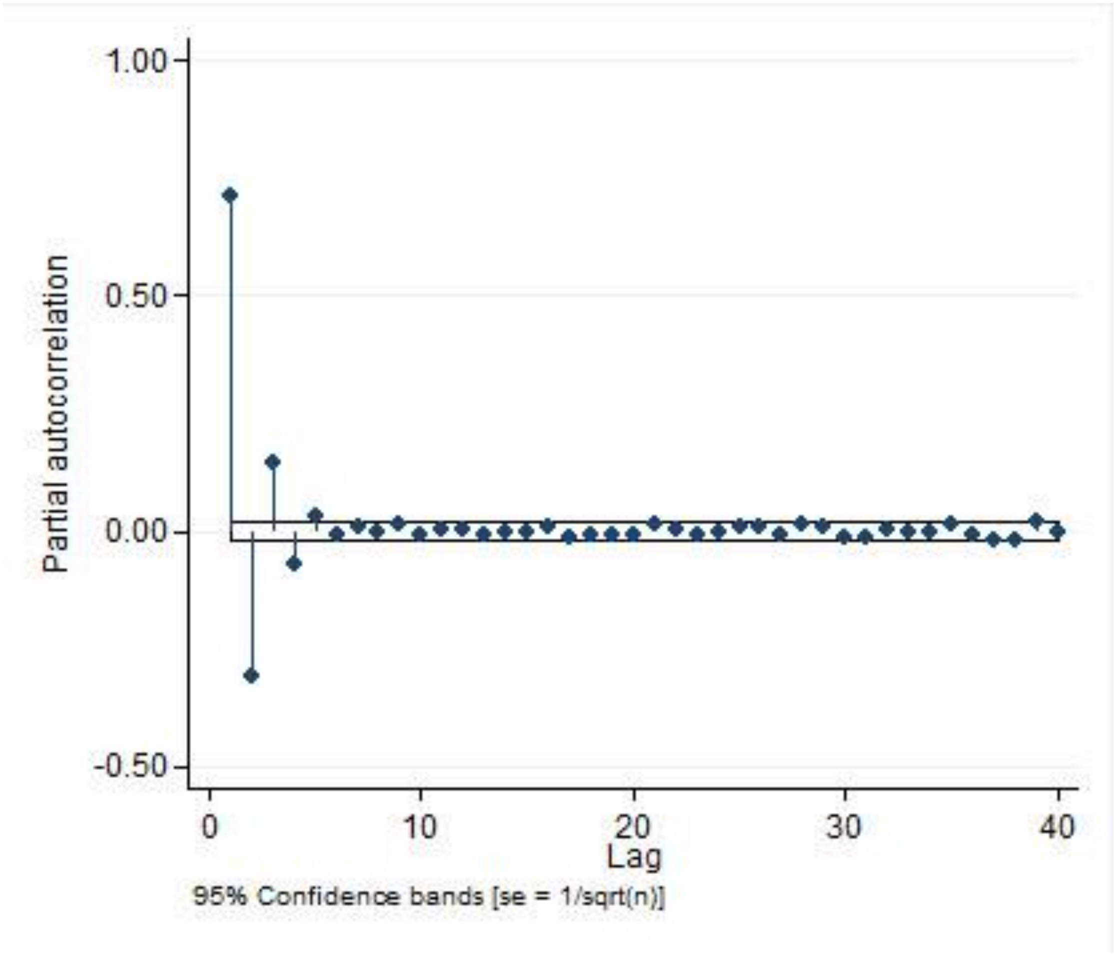
$$\Phi_p(B) X_t = \Theta_q(B) u_t$$

$$\Phi_p(B) X_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t$$

$$\Theta_q(B) u_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^q) u_t$$

Time series analysis – ARMA(1, 1) process





ACF and PACF of a ARMA(1, 1) process with p=q=0.5 (10,000 simulated values)

Time series analysis – ARIMA(p, d, q) process

Autoregressive integrated moving average process of order p, q with degree of differencing d $(\boldsymbol{X_t})_{t \in T}$

$$\Phi_p(B)(1-B)^d X_t = \Theta_q(B) u_t$$

ARIMA(p, 1, q)

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(X_t - X_{t-1}) = (1+\theta_1 B+\theta_2 B^2+\cdots+\theta_p B^q)u_t$$

$$\left(1-(\phi_1 - 1)B-(\phi_2 - \phi_1)B^2-\cdots-(\phi_p - \phi_{p-1})B^p\right)X_t$$
$$= \left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_p B^q\right)u_t$$

ARIMA(1, 1, 1)

$$(1-(\phi_1 - 1)B+\phi_1 B^2)X_t = (1+\theta_1 B)u_t$$
$$X_t = (\phi_1 - 1)X_{t-1} - \phi_1 X_{t-2} + u_t + \theta_1 u_{t-1}$$

Time series analysis – ARFIMA(p, d, q) process

Autoregressive fractionally integrated moving average process of order p, q with degree of differencing d ($-0.5 < d < 0.5$) $(X_t)_{t \in T}$

$$\Phi_p(B)(1-B)^d X_t = \Theta_q(B) u_t$$

Lag operator

$$(1-B)^d = \sum_{k=1}^{\infty} \binom{d}{k} (-B)^k$$

$$(1-B)^d = 1 - dB - \frac{1}{2} d(1-d) B^2 - \frac{1}{6} d(1-d)(2-d) B^3 - \cdots$$

Hurst exponent H

$$H = d + \frac{1}{2}$$

(Harold Edwin Hurst (1880 – 1978)
British hydrologist examining fluctuations of the water level in the Nile River)

Time series analysis – ‘long memory’

$\mathrm{d} = 0 \quad (\text{resp. } \mathrm{H} = 0.5)$

‘short memory’

Stationary and invertible ARMA processes

$-0.5 < \mathrm{d} < 0 \quad (\text{resp. } 0 < \mathrm{H} < 0.5)$

‘intermediate memory’

long-range negative dependence

Time series analysis – 'long memory'

$0 \leq d < 0.5 \quad (\text{resp. } 0.5 < H < 1)$

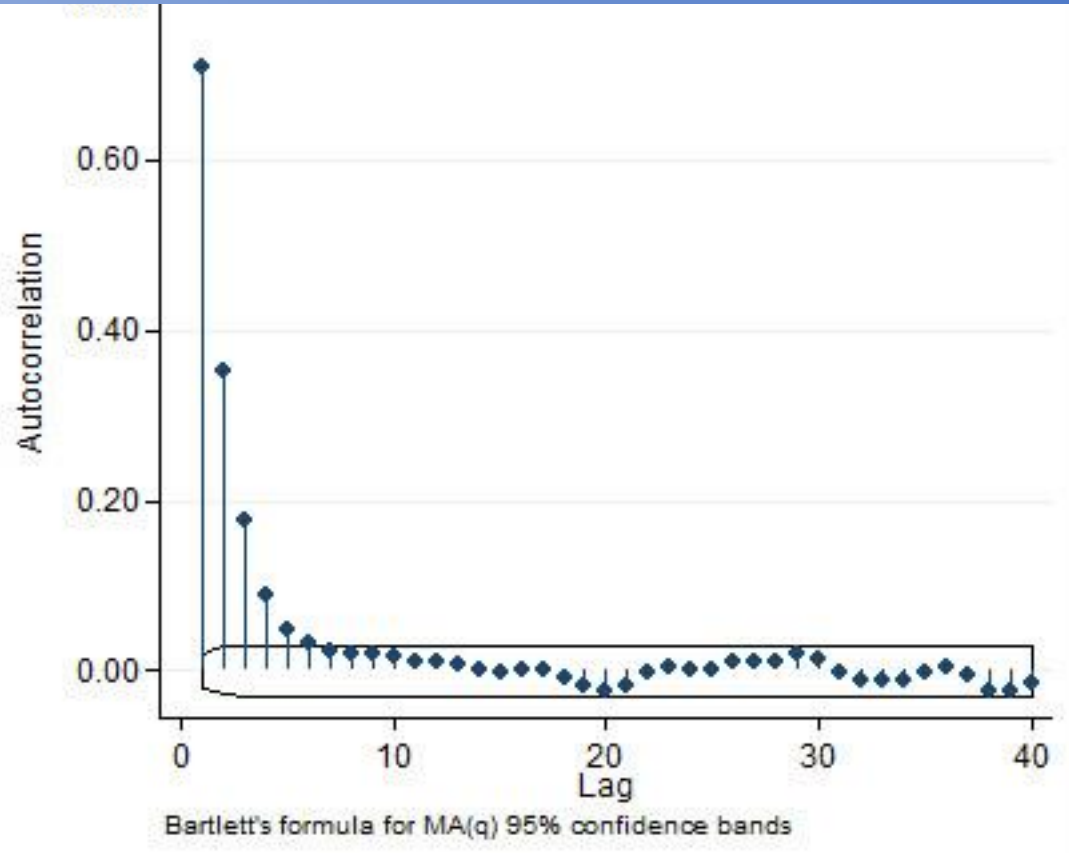
Autocorrelations of the ARFIMA processes fall hyperbolically to 0 –
in contrast to a faster, geometric decay of a stationary ARMA Process

$0 < d < 0.5 \quad (\text{resp. } 0.5 < H < 1)$

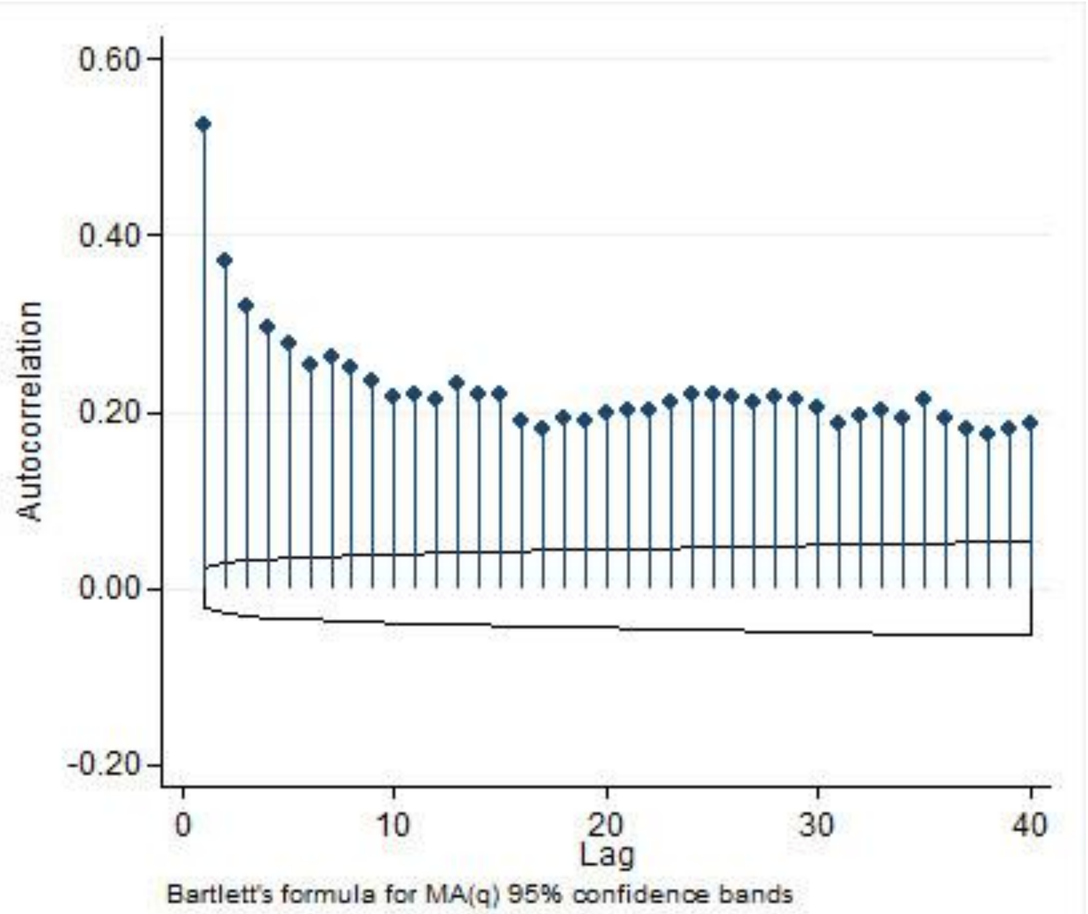
ARFIMA process shows **'long memory'**
long-range positive dependence

Possibility of long-term predictability

Time series analysis – ‘long memory’



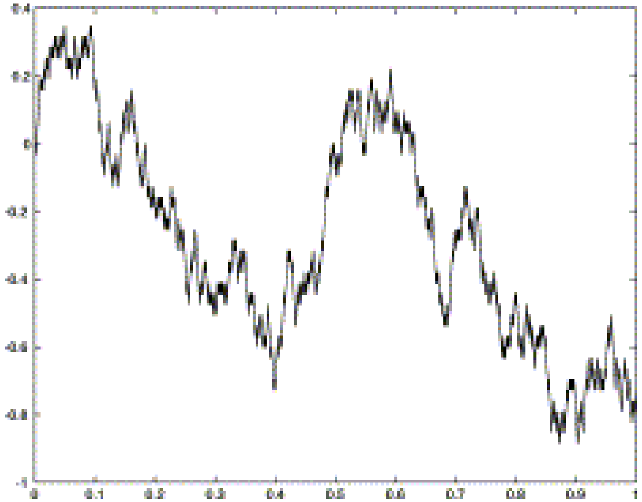
ARMA(1, 1) process with p=q=0.5

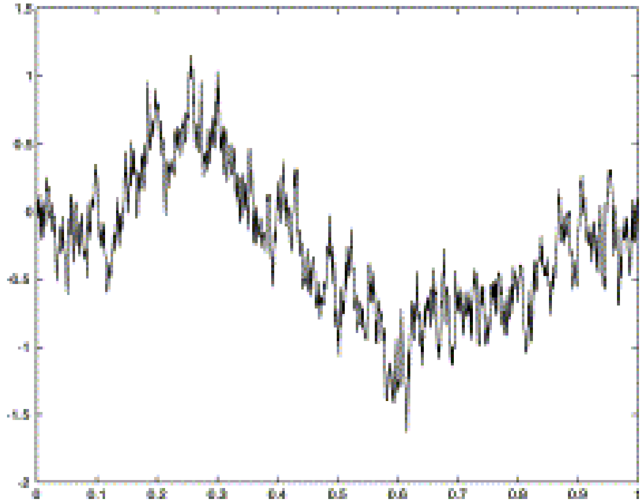


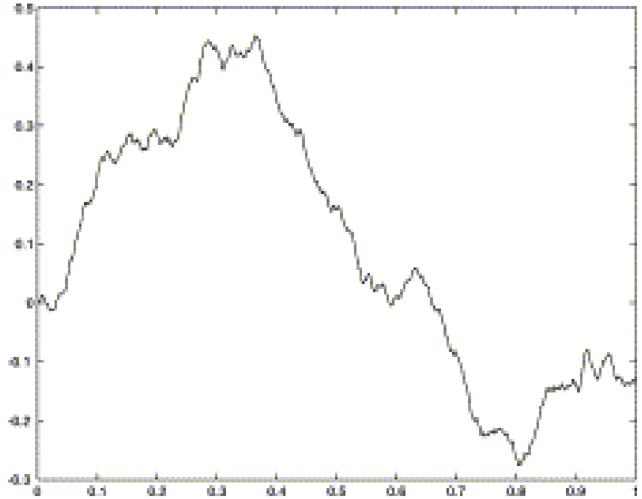
Average daily wind speed (Oberstdorf)

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Time series analysis – ‘long memory’







d=0.0 – H=0.5
‘short memory’

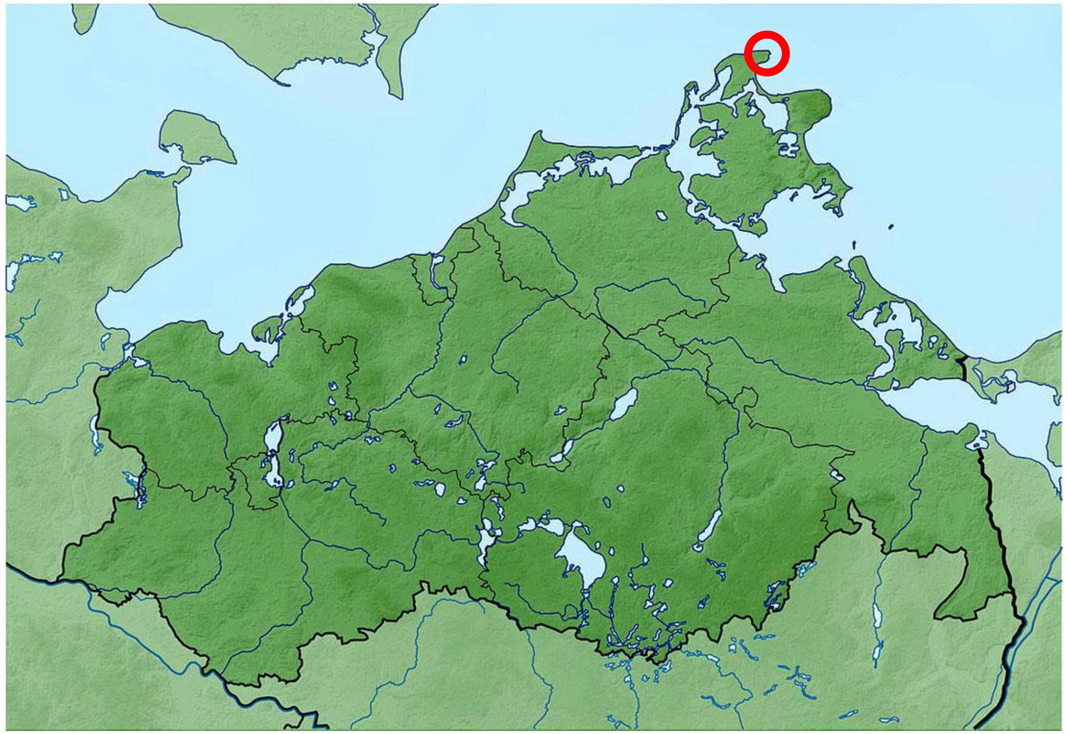
d=-0.25 – H=0.25
long-range
negative dependence

d=0.25 – H=0.75
long-range
positive dependene
‘long memory’

Source: Enriquez, N (2004). A simple construction of fractional Brownian motion. *Stochastic Processes and their Applications. 109*, 203-223

Time series analysis – seasonal decomposition

Example: Weather (wind) data Arkona



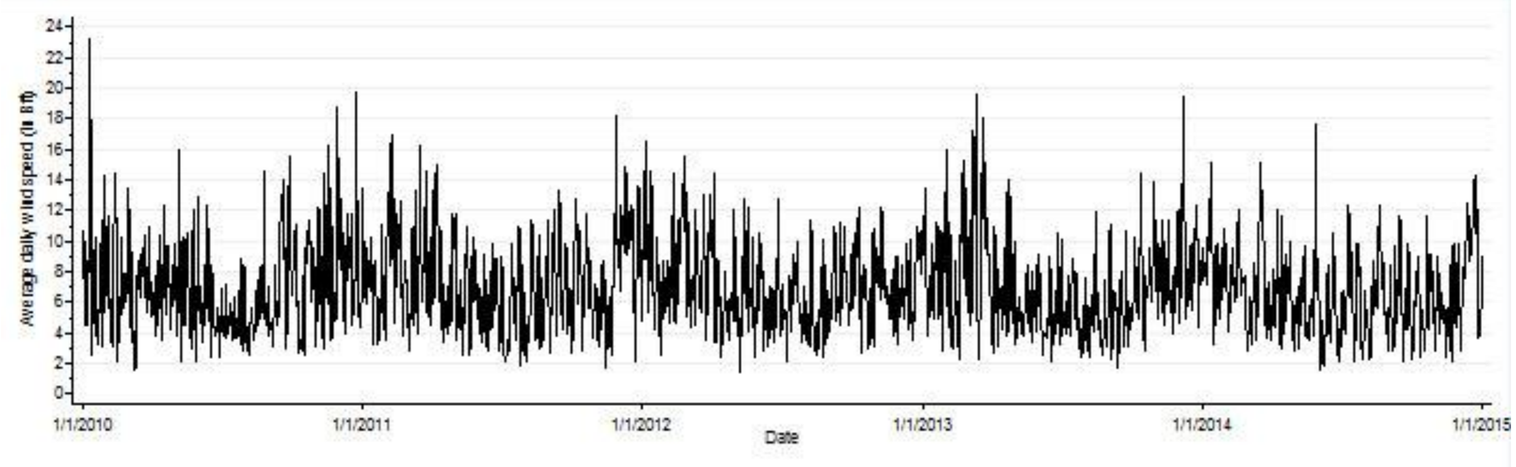
Source: https://de.wikipedia.org/wiki/Datei:Mecklenburg-Vorpommern_relief_location_map.jpg and .../Kap_Arkona#/media/File:Kap_Arkona_2012_edit.jpg

Time series analysis – seasonal decomposition

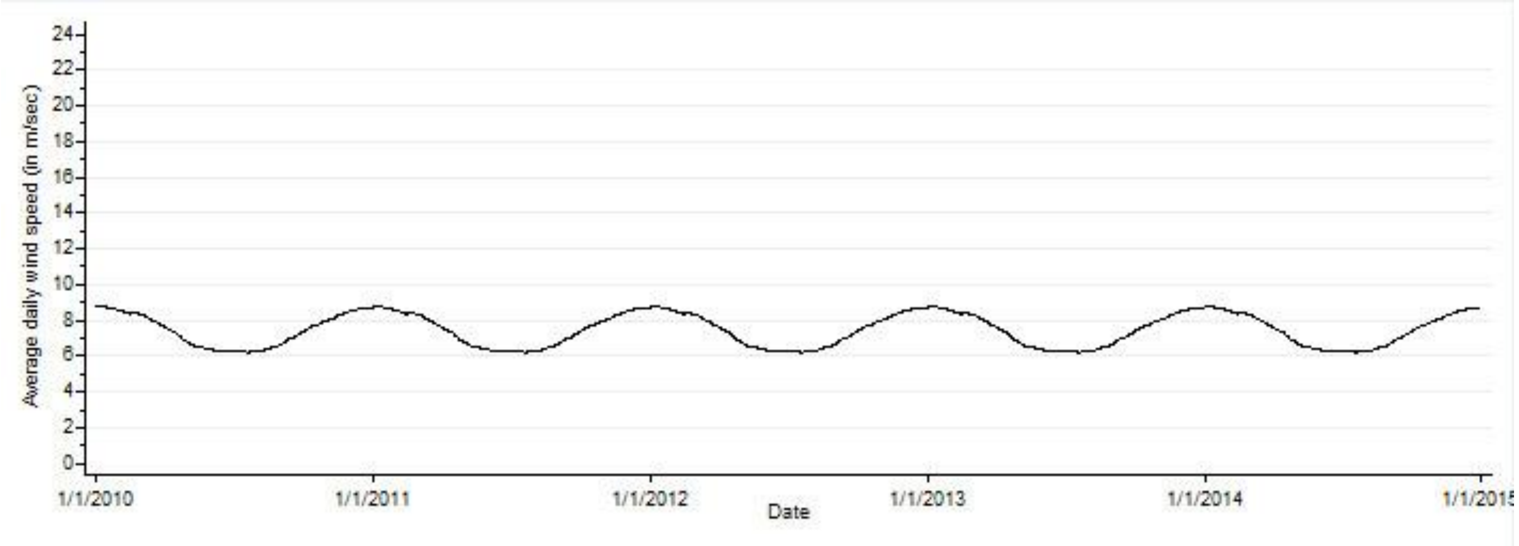
Original time series

Daily values (daily average)

Data available since 01/01/1973



Seasonal component



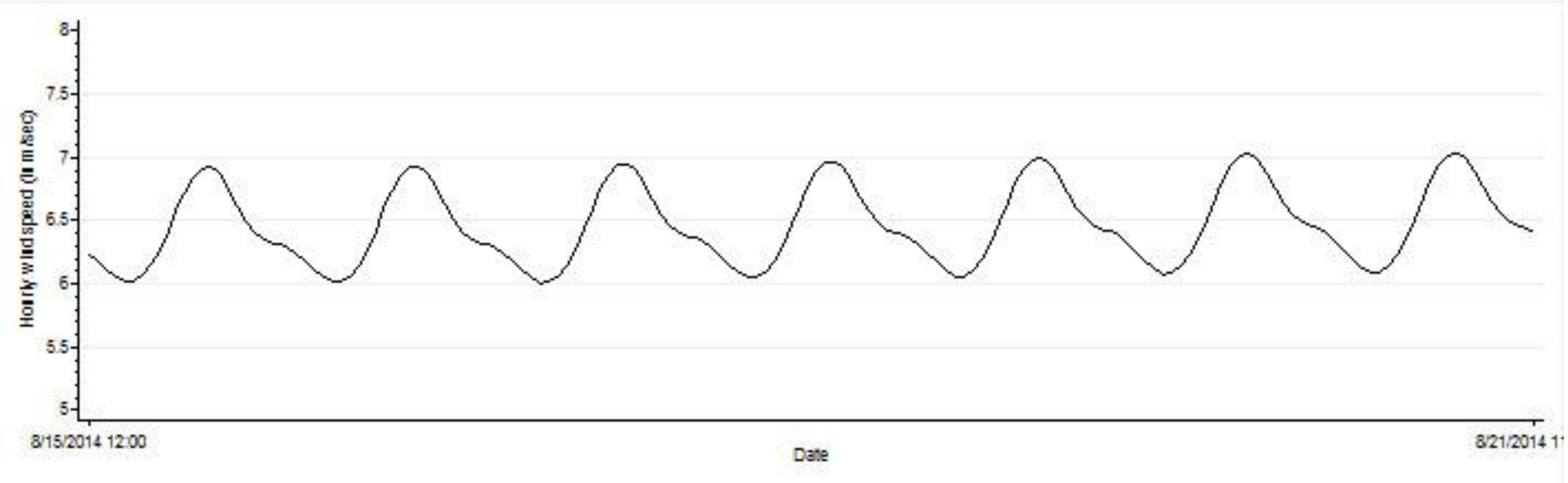
Data: DWD Deutscher Wetterdienst http://www.dwd.de

Time series analysis – seasonal decomposition using Stata

Stata	Filter used for decomposition
`tsfilter bk`	Baxter-King time-series filter
`tsfilter bw`	Butterworth time-series filter
`tsfilter cf`	Christiano-Fitzgerald time-series filter
`tsfilter hp`	Hodrick-Prescott time-series filter
`tssmooth ma`	Moving-average filter

Time series analysis – seasonal decomposition

Problems



Seasonal component

Wind

Hourly values

Data available since 01/01/1991

Multiple cycles (yearly – weekly – daily)

Very long cycles (1 year → days 8,760 hours)

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Time series analysis using ARFIMA

Background	Time series analysis
Stata ARFIMA	Results

Stata ARFIMA – sample output

Example for ARFIMA(1, 0.11, 1)

```
Sample: 1 - 7941                                Number of obs     =      7,941
                                                Wald chi2(3)      =    2297.57
Log likelihood = -19094.825                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------------
                   |                 OIM
wind_mean_woseason |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
wind_mean_woseason |
             _cons |  -.0237273   .1228598    -0.19   0.847    -.2645281    .2170735
-------------------+----------------------------------------------------------------
ARFIMA     ar   L1. |   .2005504   .0405011     4.95   0.000     .1211698     .279931
           ma   L1. |   .2149129    .029416     7.31   0.000     .1572586    .2725673
                 d |   .1130937   .0178083     6.35   0.000     .0781902    .1479973
-------------------+----------------------------------------------------------------
           /sigma2 |   7.179613   .1139405    63.01   0.000     6.956294    7.402932
------------------------------------------------------------------------------------
```

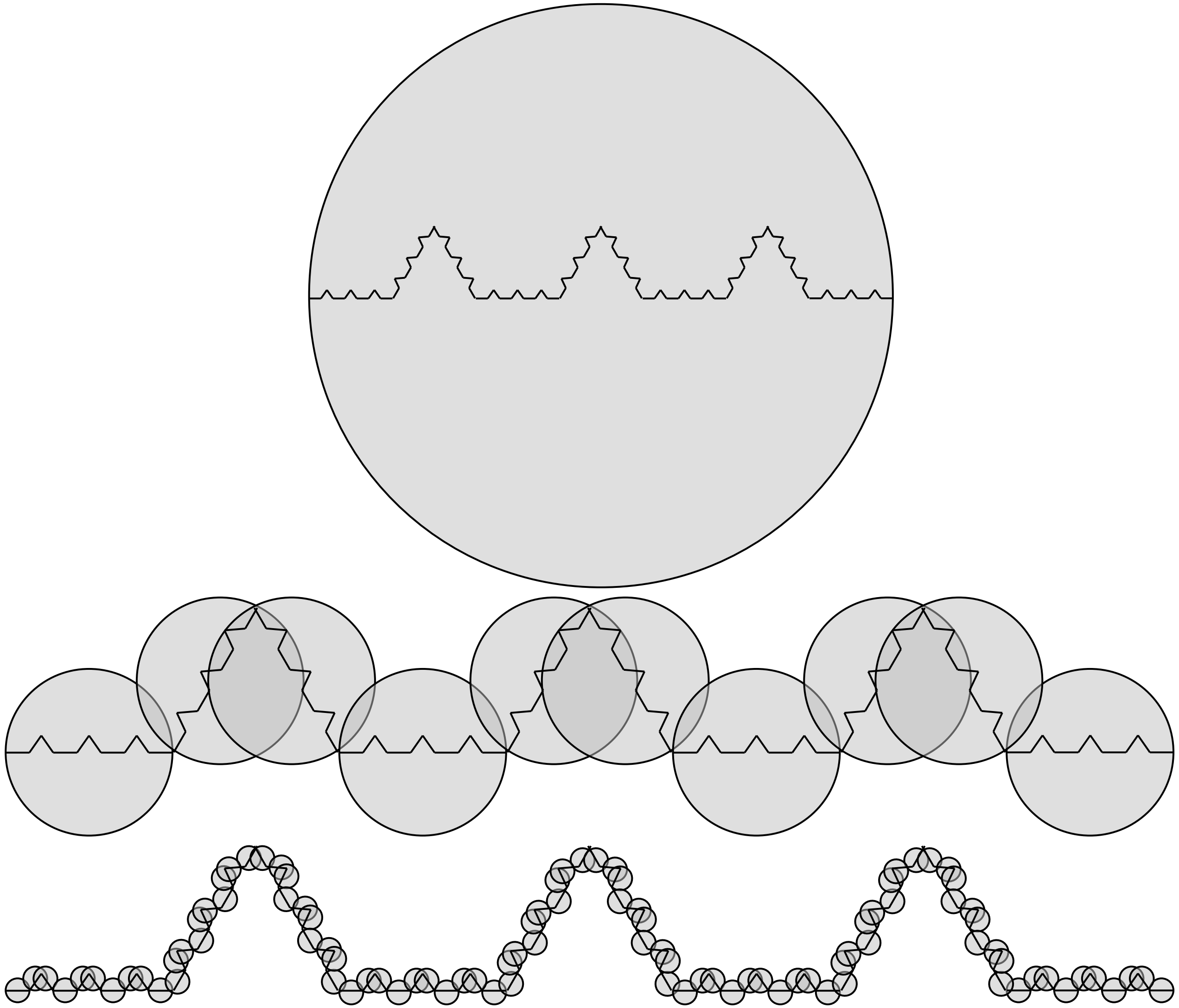
Stata ARFIMA – estimation of fractal parameter

Example:

Estimation of
Hausdorff dimension
For a version of the
Von Koch curve

More generally:

Estimation of
Scaling factors/exponents
like the
Hurst exponent



Stata ARFIMA – confidence intervals

Example for ARFIMA(1, 0.11, 1)

```
Sample: 1 - 7941                                Number of obs     =      7,941
                                                Wald chi2(3)      =    2297.57
Log likelihood = -19094.825                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------------
                   |                 OIM
wind_mean_woseason |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
wind_mean_woseason |
             _cons |  -.0237273    .1228598    -0.19   0.847    -.2645281    .2170735
-------------------+----------------------------------------------------------------
ARFIMA     ar   L1. |   .2005504    .0405011     4.95   0.000     .1211698     .279931
           ma   L1. |   .2149129     .029416     7.31   0.000     .1572586    .2725673
                  d |   .1130937    .0178083     6.35   0.000     .0781902    .1479973
-------------------+----------------------------------------------------------------
           /sigma2 |   7.179613    .1139405    63.01   0.000     6.956294    7.402932
------------------------------------------------------------------------------------
```

Stata ARFIMA – performance

Specification of starting values for the model parameters

```
matrix e_arma = 0, .2005504, .2149129, .1130937, 7.1796129
arfima wind_mean_woseason, ar(1) ma(1) iterate(50) from(e_arma)
```

Example

Step 1: ARMA(1, 1) model

Step 2: ARFIMA(1, d, 1) model

```
matrix e_arma = constant, ar(1), ma(1), 0, sigma2
```

Or

Step 1: ARFIMA(1, d, 0) model

Step 2: ARFIMA(1, d, 1) model

```
matrix e_arma = constant, ar(1) , 0, d, sigma2
```

Time series analysis using ARFIMA

Background

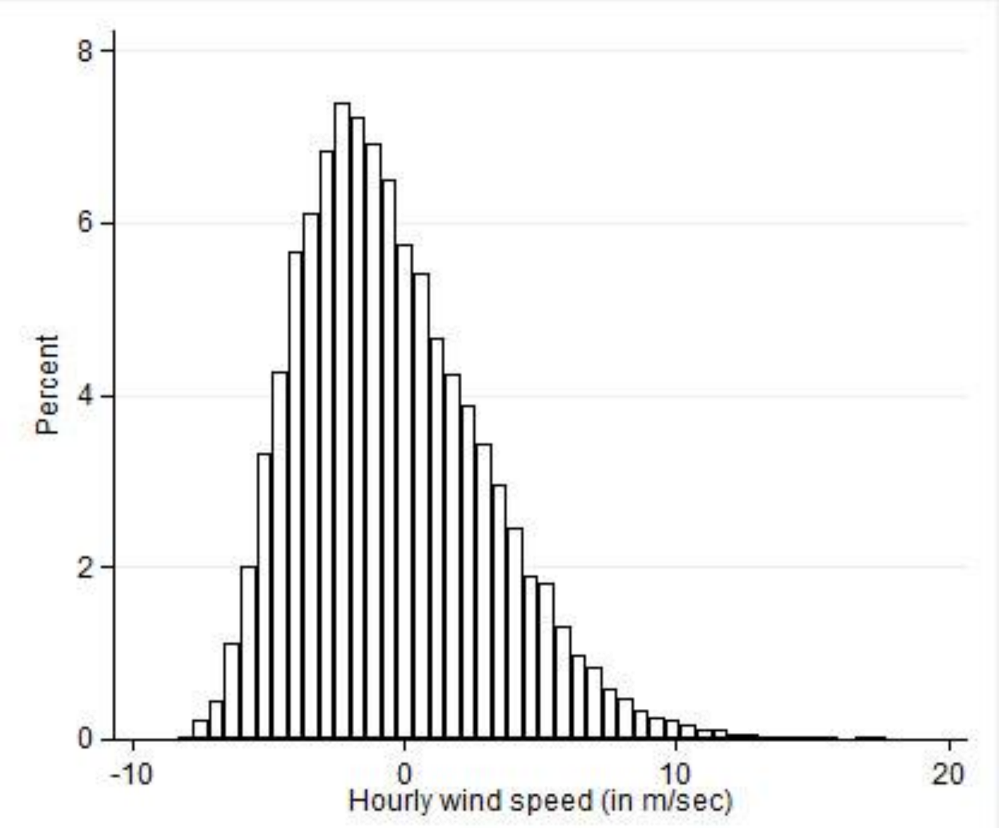
Time series analysis

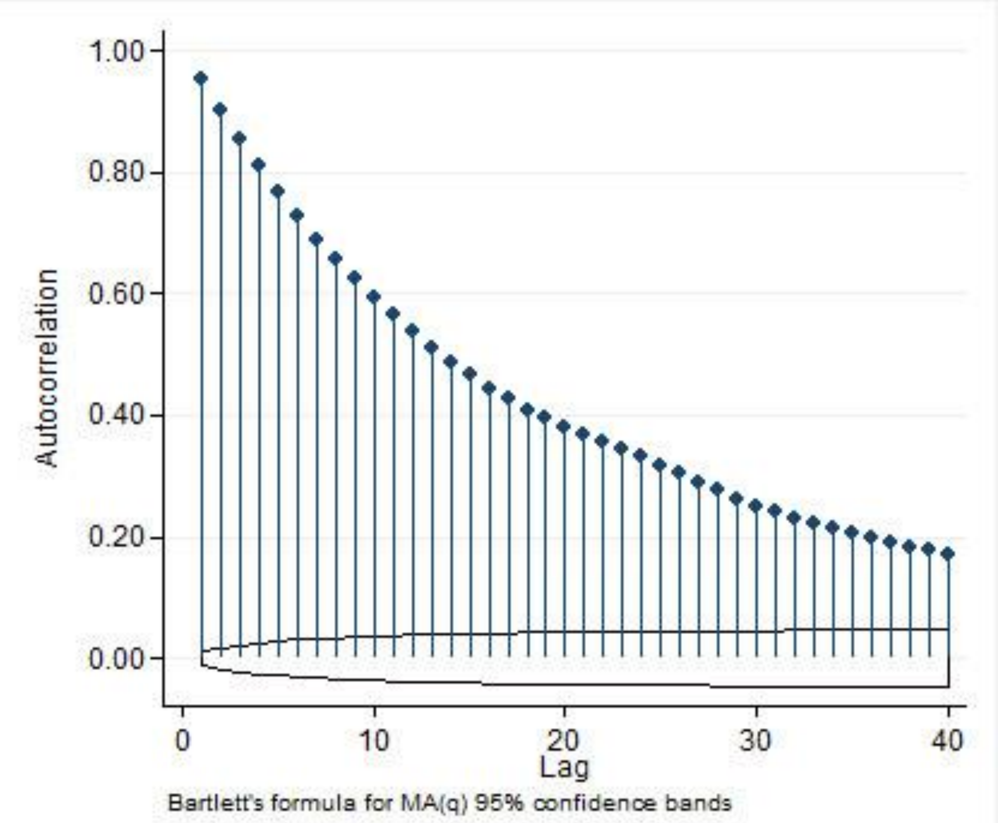
Stata ARFIMA

Results

Results – weather – wind (hourly values)

Wind – hourly values (after seasonal decomposition)
Data available since 01/01/1991 (4,192 missing values)
Used for analysis: values since 01/01/2011 (52 missing values)





Data: DWD Deutscher Wetterdienst http://www.dwd.de

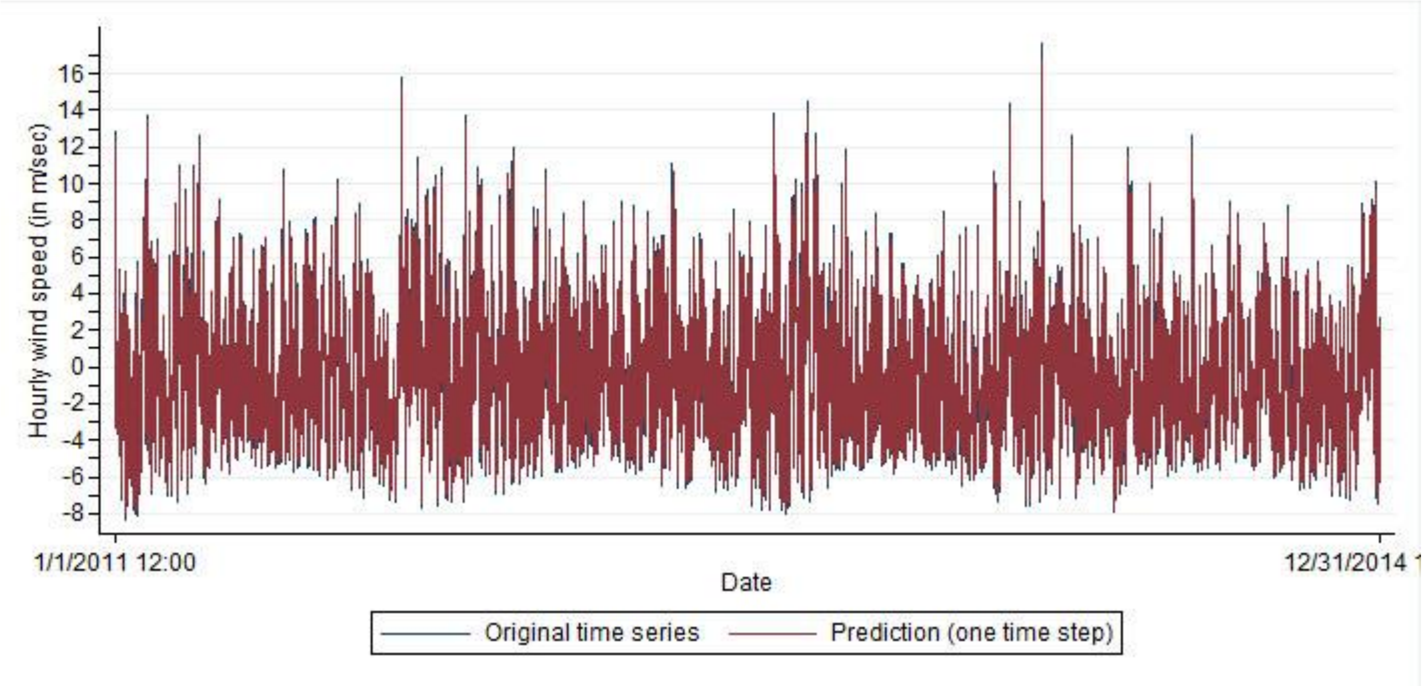
Results – weather – wind (hourly values)

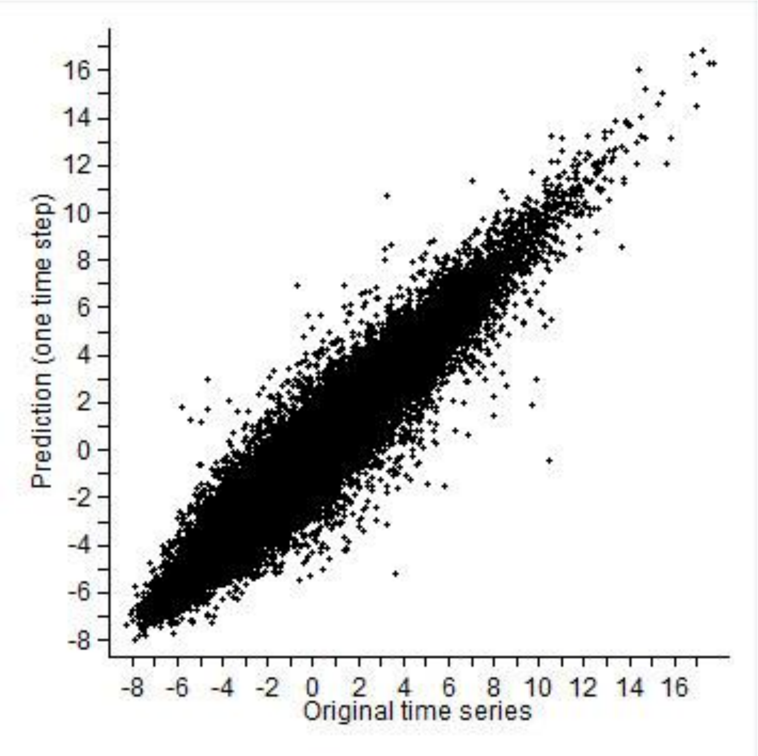
Best models	ARMA(3, 4)	ARFIMA(1, 0.14, 2)
	AR(1) = 2.06	AR(1) = 0.91
	AR(2) = -1.66	
	AR(3) = 0.33	
	AR(4) = 0.24	
	MA(1) = -1.00	
	MA(2) = 0.47	MA(2) = -0.07
	MA(3) = 0.36	
		d = 0.14

Reduction of lag terms

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Results – weather – wind (hourly values)





Overlap of prediction with original time series

Correlation 0.9555

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Results – weather – wind (hourly values)

H = d + ½	Time series
0.50	Red noise (Brownian noise, $1/f^2$ noise) → Brownian motion
0.57 … 0.64	Wind, hourly values, after seasonal decomposition
0.61 … 0.74	Wind, daily values, after seasonal decomposition
0.75	Wind, hourly values, raw data
0.73 … 0.80	Wind, daily values, raw data

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Results – weather

H = d + ½	Time series
0.50	Red noise (Brownian noise, $1/f^2$ noise) → Brownian motion
0.58	Cloudiness, average daily values, after seasonal decomposition (0 – clear / sunny ... 8 – cloudy / overcast)
0.60 ... 0.64	Relative sunshine duration, daily values, after seasonal decomposition (0 ... 100 % of timespan between astronomical sunrise and sundown)
0.77 ... 0.80	Temperature, average daily values, after seasonal decomposition
0.88	Cloudiness, average daily values, raw data

Data: DWD Deutscher Wetterdienst http://www.dwd.de

Results – energy

H = d + ½	**Time series**
0.50	Red noise (Brownian noise, $1/f^2$ noise) → Brownian motion
0.93 … 0.98	Electricity (day-ahead), daily values, after seasonal decomposition
0.95 … 0.98	Crude oil (Brent Crude), daily values, raw data

Data: https://www.epexspot.com/en/market-data/dayaheadauction http://www.ariva.de/oelpreis-brent_crude-kurs

Results – stock market

High frequency trading

DAX companies with more than 3,000 transactions per day per company

DAX 30 (German stock index)

REX / REXP (German bond market index)

ETFs (Exchange Traded Funds)

Mostly tracking of an index

Results – behavioural data

Car driving

kaggle.com
Platform for hosting public data science challenges

Data from 50,000 trips from 2,736 drivers (4.5 GByte in CSV files)

H = d + ½	Example time series
0.68	Distance between two time steps
0.84	Radial movement (Change of direction between two times steps multiplied by distance)

Thank you for your attention

Frank Ebert ebert@ebert-beratung.de

Thanks for helpful discussions:

Prof. Dr. Dr. Wilfried Grecksch, Martin-Luther-Universität Halle-Wittenberg

Dr. Steffen Rothe, Energieunion AG, Schwerin