

Quantitative Macroeconomics

Winter 2023/24

Week 2

Willi Mutschler
willi@mutschler.eu

Version: 1.0
Latest version available on: [GitHub](#)

Contents

1. Visualizing Time Series Data	1
2. Definition and Frequencies of Time Series Data	2
3. Some Fundamental Concepts Of Univariate Time Series Analysis	4
A. Solutions	6

1. Visualizing Time Series Data

1. What sort of variables are we dealing with in macroeconomics?
2. List sources and providers to get macroeconomic data.
3. Establish a strategy for obtaining macroeconomic data.
4. Analyze the quarterly growth rate of gross domestic product (GDP) for Norway. To this end, do the following:
 - Find a source for this data and download it to your computer (preferably as csv or xlsx).
 - Load the data using MATLAB's interactive import tool. Save the dates and data in two separate variables.
 - Create a new figure and plot the quarterly growth rate. Name it "Plot of Q-on-Q growth in Norway".
 - Create a new figure and plot two histograms, one with 10 bins and another one with 30 bins. Call it "Histogram of Q-on-Q growth in Norway". Compare them to a fitted normal distribution.
 - Create a new figure and assess the fit of your data to a normal distribution via a "Q-Q plot". Call it "Normal probability plot".
 - Create a new figure and plot a boxplot of your data. Call it "Boxplot".
 - Give a numerical estimate for the average growth rate and its standard deviation.
5. Redo the previous exercise for (i) 1980Q1 to 2012Q4, (ii) 2012Q4 to 2019Q4, and (iii) 2020Q1 to 2022Q2
6. Redo the whole exercise for another country of your choice. Do a Pull Request on the course's GitHub repository. to upload the used data file (preferably as csv) and your MATLAB script.

Readings

- https://de.mathworks.com/help/matlab/import_export/import-data-interactively.html
- https://de.mathworks.com/help/matlab/import_export/select-spreadsheet-data-interactively.html
- Bjørnland and Thorsrud (2015, Ch.2)

Useful resources

- Economic Data Resources (<https://libguides.umn.edu/c.php?g=843682&p=6527336>)
- Gould Library - The Data Search (<https://gouldguides.carleton.edu/c.php?g=147179&p=965273>)
- DBNomics Providers (<https://db.nomics.world/providers>)
- Our World in Data (<https://ourworldindata.org>)
- FRED (<https://fred.stlouisfed.org>)

2. Definition and Frequencies of Time Series Data

1. Briefly define a time series in terms of random variables and stochastic processes.

2. Consider data for various time series given in

- NorwayGDP.xls
- NorwayInterestRate3m.xls
- NorwayInterestRate10yrs.xls
- NorwayOSEBXGR.csv
- NorwayPopulation.xls
- NorwayRealHousePrices.xlsx
- NorwayUnemploymentRate.xls

Open the individual files and make note of the structure and source of the data. Import the data and replicate figure 1.

3. What are the data frequencies for each time series? For what kind of economic analysis would you use these frequencies?

4. What are the sample sizes? From an economic and/or statistical point of view, is it always better to have a larger sample size?

5. Roughly speaking, a time series consists of four components: a trend, a cycle, a season, and noise. To what extent do you find these features in figure 1?

6. Consider the plots in figure 1 jointly. What are possible macroeconomic issues that could be analyzed?

7. How do you aggregate time series of stock variables (like capital or debt) and of flow variables (like GDP)? For example, if you have monthly data, how do you get a quarterly time series?

Readings

- Bjørnland and Thorsrud (2015, Ch.1, Ch.2)

Various time series for Norway

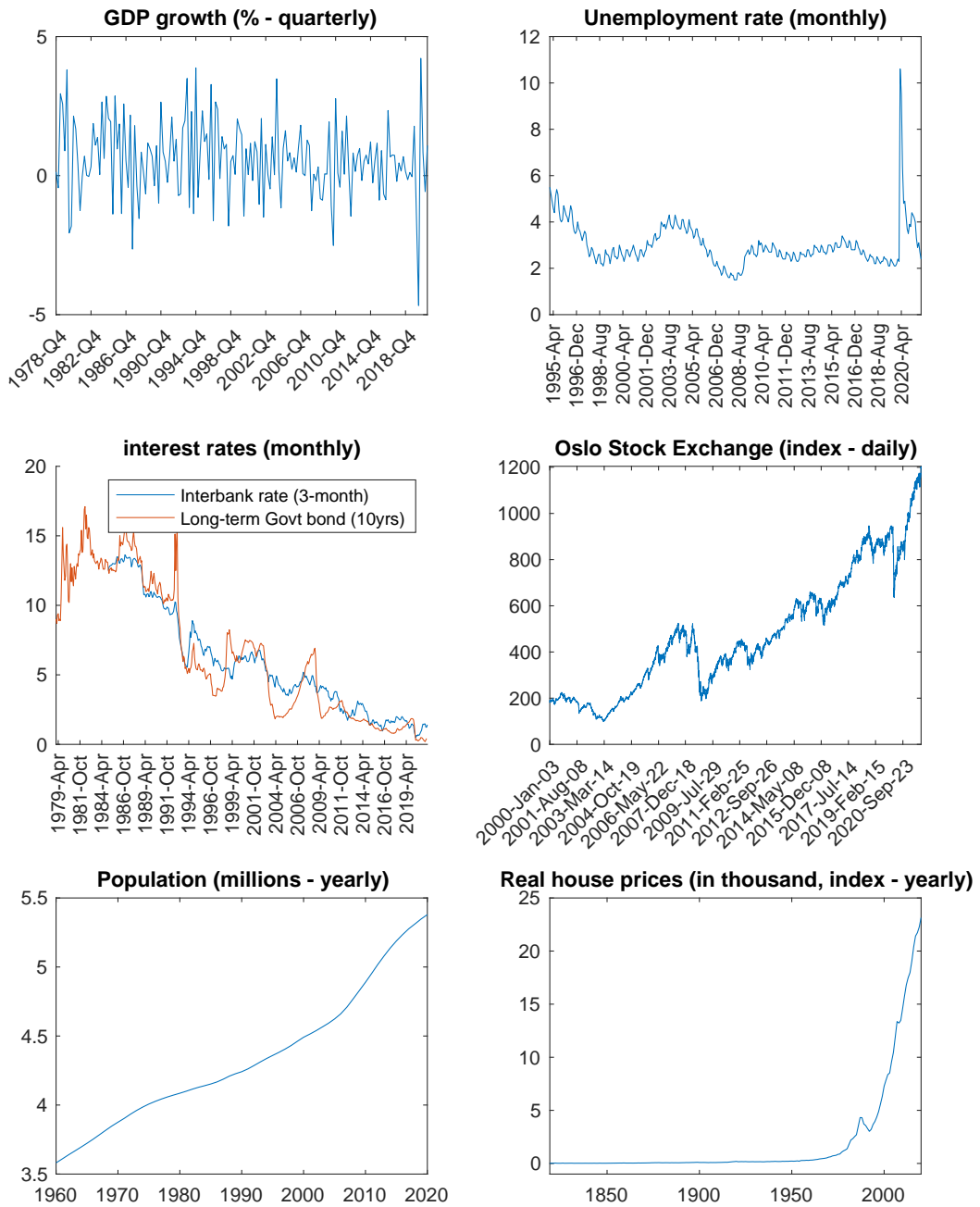


Figure 1: Various Time Series For Norway

3. Some Fundamental Concepts Of Univariate Time Series Analysis

1. Define the “White Noise Process” labeled shortly as $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.
2. Plot 200 observations of

$$(i) y_t = \varepsilon_t$$

$$(ii) y_t = \frac{1}{5}(\varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1} + \varepsilon_{t+2})$$

with $\varepsilon_t \sim N(0, 1)$. What are the differences?

3. Briefly explain the concepts of (i) weak stationarity and (ii) strict stationarity.
4. Define the autocovariance and autocorrelation function for a covariance-stationary stochastic process $\{Y_t\}$.
5. Consider the linear first-order difference equation

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

with $\varepsilon_t \sim N(0, 1)$. Simulate and plot 200 observations of (i) $|\phi| < 1$, (ii) $\phi = 1$, and (iii) $|\phi| > 1$. What does this imply in terms of stationarity of the process?

6. Briefly explain the Lag-operator and Lag-polynomials. How can we check whether an AR(p) process

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = \varepsilon_t$$

is weakly stationary?

Readings

- Bjørnland and Thorsrud (2015, Ch.2)
- Lütkepohl (2004)

References

- Ascari, Guido, Giorgio Fagiolo, and Andrea Roventini (2015). “Fat-Tail Distributions and Business-Cycle Models”. In: *Macroeconomic Dynamics* 19.2, pp. 465–476. DOI: 10.1017/S1365100513000473.
- Bjørnland, Hilde Christiane and Leif Anders Thorsrud (2015). *Applied Time Series for Macroeconomics*. 2. utgave, 1. opplag. Oslo: Gyldendal Akademisk. ISBN: 978-82-05-48089-6.
- Fagiolo, Giorgio, Mauro Napoletano, and Andrea Roventini (2008). “Are Output Growth-Rate Distributions Fat-Tailed? Some Evidence from OECD Countries”. In: *Journal of Applied Econometrics* 23.5, pp. 639–669. DOI: 10.1002/jae.1003.
- Lütkepohl, Helmut (2004). “Univariate Time Series Analysis”. In: *Applied Time Series Econometrics*. Ed. by Helmut Lütkepohl and Markus Krätzig. 1st ed. Cambridge University Press, pp. 8–85. ISBN: 978-0-521-83919-8 978-0-521-54787-1 978-0-511-60688-5. URL: <https://doi.org/10.1017/CB09780511606885.003>.

A. Solutions

1 Solution to Visualizing Time Series Data

1. Variables: gross domestic product (GDP), personal income, corporate profits, government spending, tax revenue, government deficit, unemployment rate, consumer price indices, interest rates, oil price, Greenhouse Gas emissions, debt, stock price indices, house prices, population, banking, debt securities, credit, global liquidity, derivatives, foreign exchange, property prices, and so on
2. Here are some databases:
 - European-centric sources:
 - Eurostat: official source for statistical data on the European Union, member states and sub-state regions
 - ECB Statistical Data Warehouse: Data on Euro area monetary policy, financial stability and the activities of the European System of Central Banks (ESCB), with aggregate series based on observations from national central banks, credit institutions and international data sources.
 - Data Europa EU (the former European Data Portal) provides access to over a million public datasets from 36 countries (European Union member states, the EEA, Switzerland and countries in the EU Neighborhood Policy programme). Data resources are indexed by the European Commission from national, regional, local and domain-specific public data providers.
 - US-centric sources:
 - FRED database: User-friendly database of U.S. and **international** time series data maintained by the Federal Reserve Bank of St. Louis
 - Bureau of Labor Statistics (BLS): publishes data on unemployment and consumer prices, as well as a host of data related to the U.S. labor force
 - Bureau of Economic Analysis (BEA): compiles extensive data on gross domestic product (GDP), personal income, and corporate profits
 - National Bureau of Economic Research (NBER): hosts data covering the U.S. economy, industry, and international trade
 - Congressional Budget Office: data on federal spending and revenue, projections of future spending and deficits, and forecasts
 - Worldwide:
 - Central or national banks maintain many macroeconomic statistics
 - World Bank maintains many large datasets across most countries, including the World Development Indicators (WDI) database and the Global Financial Development Database (GFDD). While they maintain an extensive set of data series with expansive country coverage, there are some missing data issues in many of the series.
 - Bank for International Settlements: datasets on international banking, debt securities, credit, global liquidity, derivatives, foreign exchange, property prices
 - United Nations maintains numerous databases, tables, and glossaries containing over 60 million data points covering international economic, health, education, and development data
 - OECD.Stat includes data and metadata for OECD countries and selected non-member economies
 - Others/Data aggregator (commercial)
 - Our World in Data

- DB NOMICS
 - Datastream and EIKON
 - Macrobond
3. It all depends on the analysis you want to conduct, i.e. whether you need country-, sector- or firm-specific data. Try to get a name for a category the dataset might belong to. Tipp: Go through the list of Providers on DBNomics to see which categories there are. Our World in Data also tends to give good ideas where to get data from. Always try official sources first, i.e. national statistical offices or central banks, then large international organizations, i.e. World Bank, OECD or BIS. Also when reading other papers, have a close look which sources have been used.
 4. The data was downloaded on October 26, 2022 from FRED and saved into a CSV (comma-separated values) file `NorwayGDP.csv`. The shortcode is `CLVMNACSCAB1GQNO`: Real Gross Domestic Product for Norway, Millions of Chained 2010 National Currency, Quarterly, Seasonally Adjusted. The time period included is 1978-01-01 to 2022-04-01. Use MATLAB's `Import Data` feature and select the columns and rows of the data. Edit the names of the variables and the input format of the dates; it is useful to tell MATLAB that the column for dates are actual dates, such that MATLAB creates a so called `datetime` array, and we can easily manipulate dates. Then click on the arrow under the green checkmark and select `Generate Script` to get MATLAB code that always will load in your data the same way. See the following script file:

`progs/matlab/visualizingTimeSeriesDataNorway.m`

```

1  % -----
2  % Visualization and descriptive statistics and plots for GDP growth in Norway
3  % -----
4  % Willi Mutschler (willi@mutschler.eu)
5  % Version: October 26, 2022
6  % -----
7
8  %% Housekeeping
9  clearvars; clc; close all;
10
11 %% Set up the Import Options and import the data
12 opts = delimitedTextImportOptions("NumVariables", 2);
13 % Specify range and delimiter
14 opts.DataLines = [2, Inf];
15 opts.Delimiter = ",";
16 % Specify column names and types
17 opts.VariableNames = ["DATE", "DATA"];
18 opts.VariableTypes = ["datetime", "double"];
19 % Specify file level properties
20 opts.ExtraColumnsRule = "ignore";
21 opts.EmptyLineRule = "read";
22 % Specify variable properties
23 opts = setvaropts(opts, "DATE", "InputFormat", "yyyy-MM-dd");
24 % Import the data; load data from different folders, e.g. '../' goes one
    subdirectory down
25 NorwayGDP = readtable("../data/NorwayGDP.csv", opts);
26 % Clear temporary variables
27 clear opts
28 NorwayGDP % note readtable has already detected dates as datetime arrays and we
    specified the format as "yyyy-MM-dd"

```

```

29 % let's change the format of the dates
30 NorwayGDP.DATE.Format = 'yyyy-QQQ'; % note that under the hood we did not really
    change the date string
31
32
33 %% Computations
34 gdp_mean = nan(1,3); %initialize variable to contain the means for different
    subsamples
35 gdp_sd = nan(1,3); %initialize variable to contain the standard deviations for
    different subsamples
36 gdp = NorwayGDP.DATA;
37 dates = NorwayGDP.DATE;
38
39 for select_sample = 1:4 % this loop runs from 1,2,3,4
40
41     if select_sample == 1 % note the double equal sign for comparisons!
42         str_sample = '1978-Q1 to 2022-Q2'; %full sample
43         sample_start = 1;
44         sample_end = length(gdp);
45     elseif select_sample == 2
46         str_sample = '1980-Q1 to 2012-Q4';
47         % dates == '01-Jan-1980' gives you a vector of 0s and 1s
48         % find() finds the position of the 1
49         sample_start = find(dates == '1980-Q1');
50         sample_end = find(dates == '2012-Q4');
51     elseif select_sample == 3
52         str_sample = '2012-Q4 to 2019-Q4';
53         sample_start = find(dates== '2012-Q4');
54         sample_end = find(dates == '2019-Q4');
55     elseif select_sample == 4
56         str_sample = '2020-Q1 to 2022-Q2';
57         sample_start = find(dates== '2020-Q1');
58         sample_end = find(dates == '2022-Q2');
59     else
60         error('select_sample needs to be 1,2,3,4');
61     end
62
63     % Compute growth rate, note that I make use of ":" and of "./"
64     gdp_growth = ( gdp((sample_start+1):sample_end) - gdp(sample_start:(
        sample_end-1)) ) ./ gdp(sample_start:(sample_end-1)); % exact
65     gdp_log_dev = log(gdp((sample_start+1):sample_end)) - log(gdp(sample_start:(
        sample_end-1))); % log approximation
66     dates_subsample = dates((sample_start+1):sample_end);
67
68     %% Make figures
69     figure('name', ['Plot of Q-on-Q growth in Norway from ', str_sample]); % this
        opens a new window and names it
70     hold on; %this enables you to draw in the same window
71     plot(dates_subsample, gdp_growth, 'linewidth', 2, 'Color', 'red', 'LineStyle', '-.');
        % simple plot with some options
72     plot(dates_subsample, gdp_log_dev, 'linewidth', 2, 'Color', 'blue', 'LineStyle', '-');
        % simple plot with some options

```

```

73 legend('Exact','Log') % this creates a legend, there are more options to it
74 title(['Plot of Q-on-Q growth in Norway from ',str_sample]); %add title
75 hold off; %turn off drawing in same windows
76
77 figure('name',['Histogram of Q-on-Q growth in Norway from ',str_sample]);
78 hold on;
79 subplot(1,2,1); % number of rows x number of columns x number of plot, so
    here first plot in left column
80 histfit(gdp_growth,10,'normal'); % 'normal' adds a fitted normal distribution
81 title('10 bins');
82 subplot(1,2,2); % number of rows x number of columns x number of plot, so
    here second plot in right column
83 histfit(gdp_growth,30,'normal');
84 title('30 bins');
85 sgtitle(['Histogram of Q-on-Q growth in Norway from ',str_sample]); % add
    common title for subplots
86 hold off;
87
88 figure('name',['Normplot of data from ',str_sample]);
89 normplot(gdp_growth);
90 title(['Normal Probability Plot of data from ',str_sample])
91
92 figure('name',['Boxplot of data from ',str_sample]);
93 boxplot(gdp_growth);
94 title(['Box Plot of data from ',str_sample])
95
96 %% Estimates
97 gdp_mean(select_sample) = mean(gdp_growth,'omitnan'); % 'omitnan' removes the
    Not-A-Number values in the computations
98 gdp_sd(select_sample) = std(gdp_growth,'omitnan');
99 fprintf('%s:The empirical mean is %.4f, the empirical standard deviation is %
    f\n',str_sample,gdp_mean(select_sample),gdp_sd(select_sample));
100 % "%s" is a placeholder for strings
101 % "%.2.f" is a placeholder for floating numbers, the .2 prints 2 numbers
    after the decimal
102 end

```

Remarks:

- A growth rate can be approximated using logs

$$\frac{Y_t - Y_{t-p}}{Y_{t-p}} \approx \log \left(\frac{Y_t}{Y_{t-p}} \right)$$

where $p = 1$ would correspond to one period lagged. If we want to compute year-on-year rates we would set $p = 4$ for quarterly data.

- Normal distribution is not a very good choice, as we have asymmetry and some probability mass in the tails of the distribution.

5. See above for the code. The distributions differ significantly depending on the subsample considered. The assumption of normality is violated; thus, we are faced with skewed distributions and have to deal with possible outliers. This is very typical for macroeconomic time series data, see e.g. Ascari, Fagiolo, and Roventini (2015) or Fagiolo, Napoletano, and Roventini (2008). However, we will see that during the course we will still keep the normality assumption; so keep this in mind for later whether this is correct or not.

6. Optional:

- Clone the repository.
- Download data for e.g. Germany and put that into `data/GermanyGDP.csv`
- Create a MATLAB script with your codes under `progs/matlab/visualizingTimeSeriesDataGermany`.
- Uncomment the line at the end of `exercises/visualizing_time_series_data_solution.inc`, so that your script is included in the compiled PDF.
- Do a pull request on GitHub.

2 Solution to Definition and Frequencies of Time Series Data

1. A time series is a collection of observations Y_t indexed by the date of each observation t . For simplicity often one denotes $t = 0, 1, 2, \dots, T$, then $\{Y_t\}_0^T = \{Y_0, Y_1, Y_2, \dots, Y_T\}$ is a sequence of random variables ordered in time (each Y_t is a random variable), which we call a stochastic process. Sometimes we rely on the concept of an infinite sample and consider $\{Y_t\}_{t=-\infty}^{\infty}$ or simply $\{Y_t\}$. A stochastic process can have many outcomes, due to its randomness, and a single outcome of a stochastic process is called a sample function or realization. A time series model assigns a joint probability distribution to the stochastic process.
2. Most of the files are downloaded from FRED except NorwayOSEBXGR.csv which is downloaded from Euronext and NorwayRealHousePrices.xlsx from the Norges Bank in October 2021.

progs/matlab/definitionFrequenciesTimeSeriesData.m

```
1 % -----
2 % This script imports and plots various time series for Norway.
3 % -----
4 % Hint:
5 % 1. Always look at the Excel/csv file first (outside MATLAB)
6 % 2. Use Matlab's "Import Data" tool, select the data you want to include.
7 %   Then click on the arrow below "Import Selection" and "Generate Script"
8 % -----
9 % Willi Mutschler (willi@mutschler.eu)
10 % Version: October 26, 2022
11 % -----
12
13 %% Housekeeping
14 clearvars; clc; close all;
15 figure('Name', 'Various time series for Norway')
16 sgtitle('Various time series for Norway');
17
18 %% GDP growth (% quarterly)
19 opts = spreadsheetImportOptions("NumVariables", 2);
20 opts.Sheet = "FRED Graph"; % Specify sheet
21 opts.DataRange = "A12:B185"; % Specify range
22 opts.VariableNames = ["observation_date", "real_gdp"]; % Specify column names
23 opts.VariableTypes = ["datetime", "double"]; % Specify column types
24 opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); %
   Specify variable properties
25 NorwayGDP = readtable("../data/NorwayGDP.xls", opts, "UseExcel", false); %
   Import the data
26 gpd_growth = 100*(log(NorwayGDP.real_gdp(4:end)) - log(NorwayGDP.real_gdp(3:end
   -1)));
27 subplot(3,2,1)
28 h = plot(NorwayGDP.observation_date(4:end), gpd_growth);
29 title('GDP growth (% - quarterly)')
30 xtickformat('yyyy-QQQ')
31 set(h.Parent, 'XTick', NorwayGDP.observation_date(4:16:end)) % get more ticks
32 sampleSize(1,1) = length(gpd_growth);
33
34 %% Unemployment rate (rate - monthly)
35 opts = spreadsheetImportOptions("NumVariables", 2);
36 opts.Sheet = "FRED Graph"; % Specify sheet
37 opts.DataRange = "A12:B332"; % Specify range
```

```

38 opts.VariableNames = ["observation_date", "unemployment_rate"]; % Specify column
    names
39 opts.VariableTypes = ["datetime", "double"]; % Specify column types
40 opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); %
    Specify variable properties
41 NorwayUnemploymentRate = readtable("../data/NorwayUnemploymentRate.xls", opts,
    "UseExcel", false); % Import the data
42
43 subplot(3,2,2)
44 h = plot(NorwayUnemploymentRate.observation_date,NorwayUnemploymentRate.
    unemployment_rate);
45 title('Unemployment rate (monthly)')
46 xtickformat('yyyy-MMM')
47 set(h.Parent, 'XTick', NorwayUnemploymentRate.observation_date(4:20:end)) % get
    more ticks
48 sampleSize(1,2) = length(NorwayUnemploymentRate.unemployment_rate);
49
50 %% interest rates (rate – monthly)
51 opts = spreadsheetImportOptions("NumVariables", 2);
52 opts.Sheet = "FRED Graph"; % Specify sheet
53 opts.DataRange = "A12:B523"; % Specify range
54 opts.VariableNames = ["observation_date", "interest_rate_3m"]; % Specify column
    names
55 opts.VariableTypes = ["datetime", "double"]; % Specify column types
56 opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); %
    Specify variable properties
57 NorwayInterestRate3m = readtable("../data/NorwayInterestRate3m.xls", opts, "
    UseExcel", false); % Import the data
58
59 opts.DataRange = "A12:B452"; % Specify range
60 opts.VariableNames = ["observation_date", "interest_rate_10yrs"]; % Specify
    column names
61 NorwayInterestRate10yrs = readtable("../data/NorwayInterestRate10yrs.xls",
    opts, "UseExcel", false); % Import the data
62
63 subplot(3,2,3)
64 hold on;
65 h = plot(NorwayInterestRate10yrs.observation_date,NorwayInterestRate10yrs.
    interest_rate_10yrs);
66 plot(NorwayInterestRate3m.observation_date,NorwayInterestRate3m.interest_rate_3m)
    ;
67 legend('Interbank rate (3-month)', 'Long-term Govt bond (10yrs)')
68 title('interest rates (monthly)')
69 xtickformat('yyyy-MMM')
70 set(h.Parent, 'XTick', NorwayInterestRate3m.observation_date(4:30:end)) % get
    more ticks
71 hold off;
72 sampleSize(1,3) = length(NorwayInterestRate3m.interest_rate_3m);
73
74 %% Oslo Stock Exchange Benchmark Index (OSEBX:IND)
75 opts = delimitedTextImportOptions("NumVariables", 3);
76 opts.DataLines = [2, Inf];

```

```

77 opts.Delimiter = ",";
78 opts.VariableNames = ["time", "OSEBXGR", "Var3"];
79 opts.SelectedVariableNames = ["time", "OSEBXGR"];
80 opts.VariableTypes = ["datetime", "double", "string"];
81 opts.ExtraColumnsRule = "ignore";
82 opts.EmptyLineRule = "read";
83 opts = setvaropts(opts, "Var3", "WhitespaceRule", "preserve");
84 opts = setvaropts(opts, "Var3", "EmptyFieldRule", "auto");
85 opts = setvaropts(opts, "time", "InputFormat", "yyyy-MM-dd HH:mm");
86 NorwayOSEBXGR = readtable("../data/NorwayOSEBXGR.csv", opts);
87
88 subplot(3,2,4)
89 h = plot(NorwayOSEBXGR.time, NorwayOSEBXGR.OSEBXGR);
90 title('Oslo Stock Exchange (index – daily)')
91 xtickformat('yyyy-MM-dd')
92 set(h.Parent, 'XTick', NorwayOSEBXGR.time(1:400:end)) % get more ticks
93 sampleSize(1,4) = length(NorwayOSEBXGR.OSEBXGR);
94
95 %% Population (millions – yearly)
96 opts = spreadsheetImportOptions("NumVariables", 2);
97 opts.Sheet = "FRED Graph"; % Specify sheet
98 opts.DataRange = "A12:B72"; % Specify range
99 opts.VariableNames = ["observation_date", "population"]; % Specify column names
100 opts.VariableTypes = ["datetime", "double"]; % Specify column types
101 opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); %
    Specify variable properties
102 NorwayPopulation = readtable("../data/NorwayPopulation.xls", opts, "UseExcel",
    false); % Import the data
103
104 subplot(3,2,5)
105 plot(NorwayPopulation.observation_date, NorwayPopulation.population./1e6)
106 title('Population (millions – yearly)')
107 xtickformat('yyyy')
108 sampleSize(1,5) = length(NorwayPopulation.population);
109
110 %% Real house prices (index – yearly)
111 opts = spreadsheetImportOptions("NumVariables", 2);
112 opts.Sheet = "Composite house price indices";
113 opts.DataRange = "A15:B216";
114 opts.VariableNames = ["Year", "Real_House_Prices"];
115 opts.VariableTypes = ["string", "double"];
116 %opts = setvaropts(opts, "Year", "InputFormat", "yyyy"); % Specify variable
    properties
117 NorwayRealHousePrices = readtable("../data/NorwayRealHousePrices.xlsx", opts,
    "UseExcel", false);
118 NorwayRealHousePrices.Year = datetime(NorwayRealHousePrices.Year, 'InputFormat', '
    yyyy');
119 subplot(3,2,6)
120 plot(NorwayRealHousePrices.Year, NorwayRealHousePrices.Real_House_Prices./1000)
121 title('Real house prices (in thousand, index – yearly)')
122 xtickformat('yyyy')
123 ylim([-1 25])

```

```

124 sampleSize(1,6) = length(NorwayRealHousePrices.Real_House_Prices);
125
126 %% Save as pdf
127 print(gcf, '../..plots/NorwayDataOverviewMatlab.pdf', '-dpdf', '-fillpage')
128
129 %% display sample sizes
130 %note that "... " enables you to continue writing the command in the next line
131 array2table(sampleSize,...
132             'VariableNames',{ 'GDP Growth (q)',...
133                               'Unemployment (m)',...
134                               'Interest Rates (m)',...
135                               'Stock Exchange (d)',...
136                               'Population (y)',...
137                               'House prices (y)'}))

```

3. Wide range of frequencies:

- Daily data: Oslo Stock Exchange
- Monthly data: Interest rates, unemployment rate
- Quarterly data: GDP growth
- Yearly data: population and real house price index

For business cycle analysis one usually focuses on monthly or quarterly data; for understanding stock returns we consider daily or monthly data; for long-run growth and wealth of nations considerations yearly data might be sufficient.

4. Sample sizes vary considerably. While quarterly data of GDP growth covers a much shorter sample than yearly house prices series, the number of observations are not that different. Higher frequencies typically mean more observations. On the one hand, many results in statistics and econometrics depend on having many observations, so the more the better. On the other hand consider for example the stock market index. We have nearly 300 business day observations every year. However, information in all these daily data does not say much about, say, the overall state of the economy. We rather have several periods: prior to mid 2003 when stock index hovered around 200, run-up to the financial crisis period from mid-2007 to mid-2008, the sharp fall during the financial crisis, the continuous recovery afterwards and then Covid. From a macroeconomic perspective we rather have a couple of “informative” periods, evident in monthly or quarterly data, all other daily observations are more or less just noise. So it is not always better to have a larger sample size if is uninformative.

5. For the different figures

- GDP growth: no trend, pronounced cyclical patterns (business cycles), no seasonality (data is seasonally adjusted), much noise
- Unemployment rate: no trend, pronounced cyclical patterns (business cycles), seasonality evident (data is not seasonally adjusted), some noise
- interest rates: downward (linear) trend, pronounced cyclical patterns in long-term yields, less so in short-term, no seasonality, moderate noise
- Oslo stock Exchange: upward (piecewise-linear) trend, cyclical patterns, no seasonality, moderate noise
- Population: upward (linear) trend, no cyclical patterns, no seasonality, no noise
- Real house prices: upward (exponential) trend, no cyclical patterns, no seasonality, no noise

6. We could analysis for instance:

- While financial crisis is clearly visible in the stock exchange, the effect on unemployment rate or real house prices is hard to detect. Real income, job opportunities and consumer confidence remained high. Why?
 - Maybe monetary policy influenced the behavior of the unemployment rate as Figure c indicates expansionary monetary policy.
 - Or Norwegian oil sector is highly productive (makes up 25% of GDP) so is this the reason why the financial crisis was cushioned?
 - Are house prices in line with their fundamentals? Is there a bubble?
7. Aggregation of higher frequencies to lower frequencies is straightforward; that is, for stock variables (such as capital or debt) we simply take the value observed, i.e. $k_t^{Q1} = k_t^{m3}$, whereas for flow variables (such as GDP) we can take the mean: $y_t^{Q1} = 1/3(y_t^{m1} + y_t^{m2} + y_t^{m3})$. Disaggregation is much more difficult and we need to use tools like interpolation or spline functions etc. → not straightforward!

3 Solution to Some Fundamental Concepts Of Univariate Time Series Analysis

1. A white noise has mean zero, a constant variance and all other second-order moments (i.e. autocovariances/autocorrelations) are zero:

$$E[\varepsilon_t] = 0$$

$$Var[\varepsilon_t] = E[\varepsilon_t^2] - E[\varepsilon_t]E[\varepsilon_t] = \sigma_\varepsilon^2$$

$$Cov(\varepsilon_t, \varepsilon_s) = E[\varepsilon_t \varepsilon_s] - E[\varepsilon_t]E[\varepsilon_s] = 0 \text{ for } s \neq t$$

2. `progs/matlab/whiteNoisePlots.m`

```

1 % -----
2 % Illustration of white noise and 5 point moving average filter
3 % -----
4 % Willi Mutschler, October 26, 2022
5 % willi@mutschler.eu
6 % -----
7 clearvars; clc; close all;
8
9 %% Generate white noise
10 sigma = 1;          % set value for standard deviation
11 T = 200;           % set value for number of observations
12 epsi = randn(T,1); % draw Tx1 vector of standard normally distributed random
    variables
13 epsi = sigma*epsi; % scale with sigma to get N(0,sigma^2) distributed random
    variables (linear transformation)
14 size(epsi)        % check size of epsi, should be Tx1
15
16 yWN = nan(T,1);    % initialize a Tx1 vector with nan (not a number)
17 for t=1:T          % the variable t runs from 1,2,...,T
18     yWN(t,1) = epsi(t);
19 end
20
21 %% Generate 5-point moving average
22 yMA5 = nan(T,1);   % initialize output vector with nan
23 for t=3:T-2        % 1) since epsi is Tx1, t cannot start at 1 as we need epsi(t
    -2) -> start at 3
24                 % 2) t cannot end at T, as we end the loop with epsi(t+2) ->
    end at T-2
25     yMA5(t) = 1/5*(epsi(t-2)+epsi(t-1)+epsi(t)+epsi(t+1)+epsi(t+2));
26 end
27
28 %% Plots
29 figure('name','White Noise Plots'); % open new window for figure
30 subplot(1,2,1);      % subplot(rows,columns, plotindex)
31 plot(yWN);           % plot white noise
32 title('White Noise'); % set title
33 subplot(1,2,2);      % subplot(rows,columns, plotindex)
34 plot(yMA5);          % plot moving-average
35 title('5-point Moving Average'); % set title

```

Every simulation is different, model can thus generate an infinite set of realizations over the period $t = 1, \dots, 200$. The processes do differ in their persistence. (i) is the white noise process,

which is not persistent. (ii) is a 5-point-moving-average, which is a linear combination of white noise processes. It is smoother and more persistent and very different from just noise. Linear combinations of white noise processes build the basis of many models in time series analysis.

3. A process is said to be **N -order weakly stationary** if all its joint moments up to order N exist and are time invariant. We are particularly interested in $N = 2$, i.e. **covariance stationarity**:

$$\begin{aligned} E[Y_t] &= \mu \text{ (constant for all } t) \\ \text{Var}[Y_t] &= E[(Y_t - \mu)(Y_t - \mu)] = \gamma_0 \text{ (constant for all } t) \\ \text{Cov}[Y_{t_1}, Y_{t_1-k}] &= E[(Y_{t_1} - \mu)(Y_{t_1-k} - \mu)] = \text{Cov}[Y_{t_2}, Y_{t_2-k}] = \gamma_k \text{ (only dependent on } k) \end{aligned}$$

That is the first two moments are not dependent on t . Particularly, the autocovariance is only dependent on the time difference k , but not on the actual point in time t .

Strict stationarity: for all k and h :

$$f(Y_t, Y_{t-1}, \dots, Y_{t-k}) = f(Y_{t-h}, Y_{t-h-1}, \dots, Y_{t-h-k})$$

That is, not only the first two moments but the whole distribution is not dependent on the point in time t , but on the time difference k .

4. Autocovariance function for a covariance-stationary process:

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)]$$

where γ_0 is the variance. Autocorrelation function:

$$\rho_k = \gamma_k / \gamma_0$$

We can estimate this by using:

$$\begin{aligned} \hat{\gamma}_k &= c_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}) \\ \hat{\rho}_k &= r_k = c_k / c_0 \end{aligned}$$

Note: In most applications we don't correct the degrees of freedom for numerical reasons (e.g. to avoid singularity of autocovariance matrices in the multivariate case), i.e. the sums are not divided by $T - k - 1$ but simply by T . For $T > 100$ this does not really matter as the expressions are very close to each other.

```

5.          progs/matlab/plotsAR1.m
1  % -----
2  % Plots of different AR(1) processes and random walks
3  % -----
4  % Willi Mutschler, October, 26 2022
5  % willi@mutschler.eu
6  % -----
7
8  %% Housekeeping
9  clearvars; clc; close all;
10
11 %% Generate and plot autoregressive processes
12 phi=[-0.8 0.4 0.9 1.01]; % different values for the phi coefficient
13 sigma=1; % value for standard deviation of white noise,
    experiment with different values

```

```

14 T=200; % value for number of observations
15 Y=nan(T,size(phi,2)); % initialize output vector with nan
16 Y(1,:)=zeros(1,size(phi,2)); % set first period to zero
17
18 for j=1:size(phi,2) % loop over coefficients
19     for t=2:T % begin loop to compute AR(1) at t=2, as there is no y(0,j), i.e.
20         you cannot index with 0
21         Y(t,j)=phi(j)*Y(t-1,j)+randn()*sigma; % Simulate time series, randn
22             simply generates one draw from N(0,1), we scale the standard deviation
23             with sigma
24     end
25 end
26
27 str_phi=["\phi=-0.8","\phi=0.4","\phi=0.9","\phi=1.01"]; % create array with
28 titles of plots, note that MATLAB can handle (some) Latex expressions
29 % note the use of double quotes which creates a string array and it is easy to
30 deal with strings of different length
31
32 figure('name','AR Plots'); % open new window for figure
33 for j=1:size(phi,2) % loop over coefficients
34     subplot(2,2,j);
35     plot(Y(:,j));
36     title(str_phi(j));
37 end
38
39 %% Generate and plot random walks
40 nRW=16; % number of Random Walks to generate
41 sigma=1; % value of standard error of white noise, experiment with
42         different values
43 T=200; % number of observations
44 Y_RW=nan(T,nRW); % initialize output vector with nan
45 Y_RW(1,:)=zeros(1,nRW); % set first period to zero
46 for j = 1:nRW
47     for t=2:T
48         Y_RW(t,j)=Y_RW(t-1,j)+randn()*sigma;
49     end
50 end
51
52 figure('name','Random Walks: y_t = y_{t-1} + /varepsilon');
53 sgtitle('Random Walks')
54 for j=1:nRW
55     subplot(4,4,j);
56     plot(Y_RW(:,j));
57 end

```

Remarks: If $|\phi| < 1$ the series returns to the mean, i.e. it is stable and stationary. If $|\phi| > 1$ then it explodes, i.e. it is unstable and not stationary. $\phi = 1$ is a so-called random walk, it is the key model when working with non-stationary models. Note that the random walk incorporates many different shapes, in macroeconomic forecasts we often want to **beat** the random walk model.

6. It is a special LINEAR operator, similar to the expectation operator, and very useful when working with time series. The operator transforms one time series into another by shifting

the observation from period t to period $t - 1$: $Ly_t = y_{t-1}$ or $L^{-1}y_t = y_{t+1}$. More general: $L^k y_t = L^{k-1} Ly_t = L^{k-1} y_{t-1} = \dots = y_{t-k}$. Convenient use:

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

We can also work with lag-polynomials:

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$$

where we call p the lag order. So:

$$\phi(L)y_t = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)y_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p}$$

To check whether an AR(p) model is covariance stationarity, we need to check whether the roots of the lag-polynomial lie outside the unit circle. That is, we treat L as a complex number $z \in \mathbb{C}$ and compute the roots of $(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p) = 0$ (using a computer in most cases).