Quantitative Macroeconomics

Winter 2022/23

Week 2

Willi Mutschler
willi@mutschler.eu

Version: 1.0.0
Latest version available on: GitHub

Contents

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)

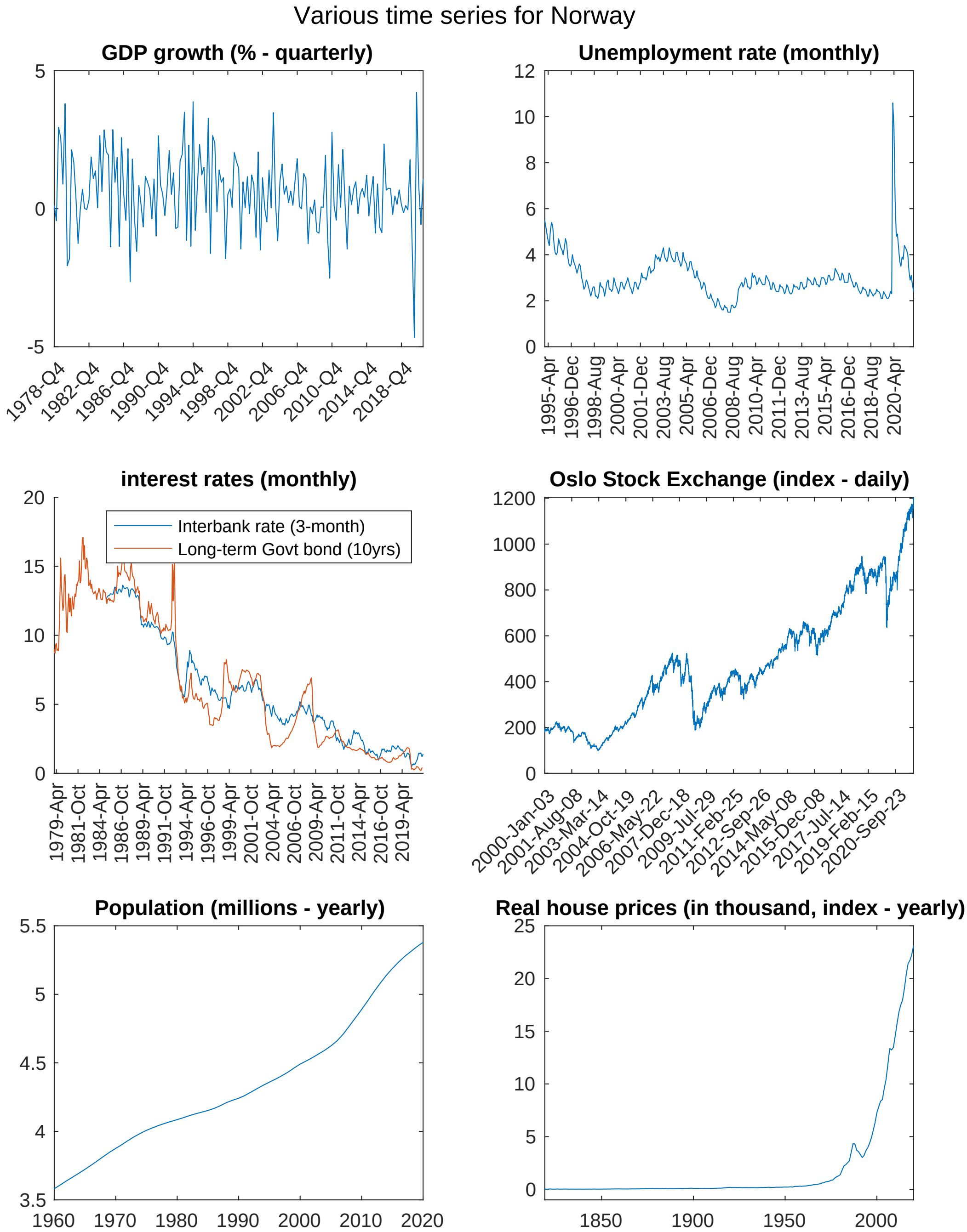

Figure 1: Various Time Series For Norway

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

$$
\begin{aligned}
(i)\ y_t &= \varepsilon_t \\
(ii)\ y_t &= \frac{1}{5}(\varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1} + \varepsilon_{t+2})
\end{aligned}
$$

   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

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

   is weakly stationary?

Readings

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

4. Properties Of AR(1)

Let $\{\varepsilon_t\}$ be a white noise process with variance σ_ε^2.

1. Consider the univariate first-order autoregressive process AR(1):

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

   Derive the conditional and unconditional first and second moments.

2. Simulate different AR(1) processes and plot the corresponding autocorrelation function (ACF). To this end write a function `ACFPlots`(y,p^{max},α) to plot the ACF of the data vector y with maximum number of lags p^{max}. Also include an approximate (1-α)% confidence interval around zero in your plot. Hints:
   - The empirical autocorrelation function at lag k is defined as $r_k = c_k/c_0$ where

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

     and

$$c_0 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})$$

   - You can either use a for-loop to compute the sum or use vectors: $(y - \bar{y})'(y - \bar{y})$.
   - The sample autocorrelation function is an estimate of the actual autocorrelation only if the process is stationary. If the process is purely random, that is, all members are mutually independent and identically distributed so that y_t and y_{t-k} are stochastically independent for $k \neq 0$, then the normalized estimated autocorrelations are asymptotically standard normally distributed, i.e. $\sqrt{T}r_k \to U \sim N(0,1)$ and thus $r_k \to \tilde{U} \sim N(0, 1/T)$.

Hints:

- If $|\phi| < 1$, then $\sum_{j=0}^{\infty}\phi^j = \frac{1}{(1-\phi)}$

Readings

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

5. Properties AR(1) With Time Trend

Consider the univariate AR(1) model with a constant and time trend

$$y_t = c + d \cdot t + \phi y_{t-1} + u_t$$

where $u_t \sim WN(0, \sigma_u^2)$, $|\phi| < 1$, $c \in \mathbb{R}$ and $d \in \mathbb{R}$.

1. Compute the unconditional first and second moments, i.e. the unconditional mean, variance, autocovariance and autocorrelation of y_t.
2. Why is this process not covariance-stationary? How could one proceed to make it covariance-stationary?

Hints:

- If $|\phi| < 1$, then $\sum_{j=0}^{\infty} \phi^j j = \frac{\phi}{(1-\phi^2)}$

Readings

- 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. First. Cambridge University Press, pp. 8–85. ISBN: 978-0-521-83919-8 978-0-521-54787-1 978-0-511-60688-5. DOI: `10.1017/CBO9780511606885.003`.

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-seperated 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:

```
% =========================================================================
% VisualizingTimeSeriesData.m
% =========================================================================
% Willi Mutschler (willi@mutschler.eu)
% Version: October 26, 2022
% =========================================================================

%% Housekeeping
clearvars; clc;close all;

%% Set up the Import Options and import the data
opts = delimitedTextImportOptions("NumVariables", 2);
% Specify range and delimiter
opts.DataLines = [2, Inf];
opts.Delimiter = ",";
% Specify column names and types
opts.VariableNames = ["DATE", "DATA"];
opts.VariableTypes = ["datetime", "double"];
% Specify file level properties
opts.ExtraColumnsRule = "ignore";
opts.EmptyLineRule = "read";
% Specify variable properties
opts = setvaropts(opts, "DATE", "InputFormat", "yyyy-MM-dd");
% Import the data; load data from different folders, e.g. '../' goes one subdirectory down
NorwayGDP = readtable("../../data/NorwayGDP.csv", opts);
% Clear temporary variables
clear opts
NorwayGDP % note readtable has already detected dates as datetime arrays and we specified the format as "yyyy-MM-dd"
% let's change the format of the dates
```

```
NorwayGDP.DATE.Format = 'yyyy-QQQ'; % note that under the hood we did not really change the date string


%% Computations
gdp_mean = nan(1,3); %initialize variable to contain the means for different subsamples
gdp_sd = nan(1,3);   %initialize variable to contain the standard deviations for different subsamples
gdp = NorwayGDP.DATA;
dates = NorwayGDP.DATE;

for select_sample = 1:4 % this loop runs from 1,2,3,4

    if select_sample == 1  % note the double equal sign for comparisons!
        str_sample = '1978-Q1 to 2022-Q2'; %full sample
        sample_start = 1;
        sample_end   = length(gdp);
    elseif select_sample == 2
        str_sample = '1980-Q1 to 2012-Q4';
        % dates == '01-Jan-1980' gives you a vector of 0s and 1s
        % find() finds the position of the 1
        sample_start = find(dates == '1980-Q1');
        sample_end   = find(dates == '2012-Q4');
    elseif select_sample == 3
        str_sample = '2012-Q4 to 2019-Q4';
        sample_start = find(dates== '2012-Q4');
        sample_end   = find(dates == '2019-Q4');
    elseif select_sample == 4
        str_sample = '2020-Q1 to 2022-Q2';
        sample_start = find(dates== '2020-Q1');
        sample_end   = find(dates == '2022-Q2');
    else
        error('select_sample needs to be 1,2,3,4');
    end

    % Compute growth rate, note that I make use of ":" and of "./"
    gdp_growth = ( gdp((sample_start+1):sample_end) - gdp(sample_start:(sample_end-1)) ) ./ gdp(sample_start:(sample_end-1)); % exact
    gdp_log_dev = log(gdp((sample_start+1):sample_end)) - log(gdp(sample_start:(sample_end-1))); % log approximation
    dates_subsample = dates((sample_start+1):sample_end);

    %% Make figures
    figure('name',['Plot of Q-on-Q growth in Norway from ',str_sample]); % this opens a new window and names it
    hold on; %this enables you to draw in the same window
    plot(dates_subsample,gdp_growth,'linewidth',2,'Color','red','LineStyle','-.');   % simple plot with some options
    plot(dates_subsample,gdp_log_dev,'linewidth',2,'Color','blue','LineStyle','--'); % simple plot with some options
    legend('Exact','Log') % this creates a legend, there are more options to it
```

```
    title(['Plot of Q-on-Q growth in Norway from',str_sample]); %add title
    hold off; %turn off drawing in same windows

    figure('name',['Histogram of Q-on-Q growth in Norway from ',str_sample]);
    hold on;
    subplot(1,2,1); % number of rows x number of columns x number of plot, so here first plot in left column
    histfit(gdp_growth,10,'normal'); % 'normal' adds a fitted normal distribution
    title('10 bins');
    subplot(1,2,2); % number of rows x number of columns x number of plot, so here second plot in right column
    histfit(gdp_growth,20,'normal');
    title('30 bins');
    sgtitle(['Histogram of Q-on-Q growth in Norway from ',str_sample]); % add common title for subplots
    hold off;

    figure('name',['Normplot of data from ',str_sample]);
    normplot(gdp_growth);
    title(['Normal Probability Plot of data from ',str_sample])

    figure('name',['Boxplot of data from ',str_sample]);
    boxplot(gdp_growth);
    title(['Box Plot of data from ',str_sample])

    %% Estimates
    gdp_mean(select_sample) = mean(gdp_growth,'omitnan'); % 'omitnan' removes the Not-A-Number values in the computations
    gdp_sd(select_sample) = std(gdp_growth,'omitnan');
    fprintf('%s:The empirical mean is %.4f, the empirical standard deviation is %f\n',str_sample,gdp_mean(select_sample),gdp_sd(select_sample));
    % "%s" is a placeholder for strings
    % "%.2.f" is a placeholder for floating numbers, the .2 prints 2 numbers after the decimal
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 $t = 0, 1, 2, ..., T$, then $\{Y_t\}_0^T = \{Y_0, Y_1, Y_2, ..., Y_T\}$ 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.

```
% =========================================================================
% DefinitionFrequenciesTimeSeriesData.m
% =========================================================================
% Hint:
% 1. Always look at the Excel/csv file first (outside MATLAB)
% 2. Use Matlab's "Import Data" tool, select the data you want to include.
%    Then click on the arrow below "Import Selection" and "Generate Script"
% =========================================================================
% Willi Mutschler (willi@mutschler.eu)
% Version: October 26, 2022
% =========================================================================

%% Housekeeping
clearvars; clc;close all;
figure('Name','Various time series for Norway')
sgtitle('Various time series for Norway');

%% GDP growth (% quarterly)
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "FRED Graph"; % Specify sheet
opts.DataRange = "A12:B185"; % Specify range
opts.VariableNames = ["observation_date", "real_gdp"]; % Specify column names
opts.VariableTypes = ["datetime", "double"]; % Specify column types
opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); % Specify variable properties
NorwayGDP = readtable("../../data/NorwayGDP.xls", opts, "UseExcel", false); % Import the data
gpd_growth = 100*(log(NorwayGDP.real_gdp(4:end)) - log(NorwayGDP.real_gdp(3:end-1)));
subplot(3,2,1)
h = plot(NorwayGDP.observation_date(4:end),gpd_growth);
title('GDP growth (% - quarterly)')
xtickformat('yyyy-QQQ')
set(h.Parent, 'XTick', NorwayGDP.observation_date(4:16:end)) % get more ticks
sampleSize(1,1) = length(gpd_growth);
%% Unemployment rate (rate - monthly)
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "FRED Graph"; % Specify sheet
opts.DataRange = "A12:B332"; % Specify range
opts.VariableNames = ["observation_date", "unemployment_rate"]; % Specify column names
```

```
opts.VariableTypes = ["datetime", "double"]; % Specify column types
opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); % Specify variable properties
NorwayUnemploymentRate = readtable("../../data/NorwayUnemploymentRate.xls", opts, "UseExcel", false); % Import the data

subplot(3,2,2)
h = plot(NorwayUnemploymentRate.observation_date,NorwayUnemploymentRate.unemployment_rate);
title('Unemployment rate (monthly)')
xtickformat('yyyy-MMM')
set(h.Parent, 'XTick', NorwayUnemploymentRate.observation_date(4:20:end)) % get more ticks
sampleSize(1,2) = length(NorwayUnemploymentRate.unemployment_rate);

%% interest rates (rate - monthly)
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "FRED Graph"; % Specify sheet
opts.DataRange = "A12:B523"; % Specify range
opts.VariableNames = ["observation_date", "interest_rate_3m"]; % Specify column names
opts.VariableTypes = ["datetime", "double"]; % Specify column types
opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); % Specify variable properties
NorwayInterestRate3m = readtable("../../data/NorwayInterestRate3m.xls", opts, "UseExcel", false); % Import the data

opts.DataRange = "A12:B452"; % Specify range
opts.VariableNames = ["observation_date", "interest_rate_10yrs"]; % Specify column names
NorwayInterestRate10yrs = readtable("../../data/NorwayInterestRate10yrs.xls", opts, "UseExcel", false); % Import the data

subplot(3,2,3)
hold on;
h = plot(NorwayInterestRate10yrs.observation_date,NorwayInterestRate10yrs.interest_rate_10yrs);
plot(NorwayInterestRate3m.observation_date,NorwayInterestRate3m.interest_rate_3m);
legend('Interbank rate (3-month)','Long-term Govt bond (10yrs)')
title('interest rates (monthly)')
xtickformat('yyyy-MMM')
set(h.Parent, 'XTick', NorwayInterestRate3m.observation_date(4:30:end)) % get more ticks
hold off;
sampleSize(1,3) = length(NorwayInterestRate3m.interest_rate_3m);

%% Oslo Stock Exchange Benchmark Index (OSEBX:IND)
opts = delimitedTextImportOptions("NumVariables", 3);
opts.DataLines = [2, Inf];
opts.Delimiter = ",";
opts.VariableNames = ["time", "OSEBXGR", "Var3"];
```

```
opts.SelectedVariableNames = ["time", "OSEBXGR"];
opts.VariableTypes = ["datetime", "double", "string"];
opts.ExtraColumnsRule = "ignore";
opts.EmptyLineRule = "read";
opts = setvaropts(opts, "Var3", "WhitespaceRule", "preserve");
opts = setvaropts(opts, "Var3", "EmptyFieldRule", "auto");
opts = setvaropts(opts, "time", "InputFormat", "yyyy-MM-dd HH:mm");
NorwayOSEBXGR = readtable("../../data/NorwayOSEBXGR.csv", opts);

subplot(3,2,4)
h = plot(NorwayOSEBXGR.time,NorwayOSEBXGR.OSEBXGR);
title('Oslo Stock Exchange (index - daily)')
xtickformat('yyyy-MMM-dd')
set(h.Parent, 'XTick', NorwayOSEBXGR.time(1:400:end)) % get more ticks
sampleSize(1,4) = length(NorwayOSEBXGR.OSEBXGR);

%% Population (millions - yearly)
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "FRED Graph"; % Specify sheet
opts.DataRange = "A12:B72"; % Specify range
opts.VariableNames = ["observation_date", "population"]; % Specify column names
opts.VariableTypes = ["datetime", "double"]; % Specify column types
opts = setvaropts(opts, "observation_date", "InputFormat", "yyyy-MM-dd"); % Specify variable properties
NorwayPopulation = readtable("../../data/NorwayPopulation.xls", opts, "UseExcel", false); % Import the data

subplot(3,2,5)
plot(NorwayPopulation.observation_date,NorwayPopulation.population./1e6)
title('Population (millions - yearly)')
xtickformat('yyyy')
sampleSize(1,5) = length(NorwayPopulation.population);

%% Real house prices (index - yearly)
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "Composite house price indices";
opts.DataRange = "A15:B216";
opts.VariableNames = ["Year", "Real_House_Prices"];
opts.VariableTypes = ["string", "double"];
%opts = setvaropts(opts, "Year", "InputFormat", "yyyy"); % Specify variable properties
NorwayRealHousePrices = readtable("../../data/NorwayRealHousePrices.xlsx", opts, "UseExcel", false);
NorwayRealHousePrices.Year = datetime(NorwayRealHousePrices.Year,'InputFormat','yyyy');
subplot(3,2,6)
plot(NorwayRealHousePrices.Year,NorwayRealHousePrices.Real_House_Prices./1000)
title('Real house prices (in thousand, index - yearly)')
xtickformat('yyyy')
ylim([-1 25])
sampleSize(1,6) = length(NorwayRealHousePrices.Real_House_Prices);

```

```
%% Save as pdf
print(gcf, '../../plots/NorwayDataOverviewMatlab.pdf', '-dpdf', '-fillpage')

%% display sample sizes
%note that "..." enables you to continue writing the command in the next line
array2table(sampleSize,...
            'VariableNames',{'GDP Growth (q)',...
                             'Unemployment (m)',...
                             'Interest Rates (m)',...
                             'Stock Exchange (d)',...
                             'Population (y)',...
                             '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. $\rightarrow$ 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:

$$
\begin{aligned}
E[\varepsilon_t] &= 0 \\
Var[\varepsilon_t] &= E[\varepsilon_t^2] - E[\varepsilon_t]E[\varepsilon_t] = \sigma_\varepsilon^2 \\
Cov(\varepsilon_t, \varepsilon_s) &= E[\varepsilon_t \varepsilon_s] - E[\varepsilon_t]E[\varepsilon_s] = 0 \text{ for } s \neq t
\end{aligned}
$$

2.

```matlab
% =========================================================================
% WhiteNoisePlots.m
% =========================================================================
% Willi Mutschler, October 26, 2022
% willi@mutschler.eu
% =========================================================================
clearvars; clc;close all;

%% Generate white noise
sigma = 1;           % set value for standard deviation
T = 200;             % set value for number of observations
epsi = randn(T,1);   % draw Tx1 vector of Gaussian random variables
size(epsi)           % check size of epsi, should be Tx1

y1 = nan(T,1);       % initialize a Tx1 vector with nan (not a number)
for t=1:T            % the variable t runs from 1,2....,T
    y1(t,1) = epsi(t)*sigma;
    % epsi are drawn from N(0,1), we scale the standard error by multiplying with sigma
end

%% Generate 5-point moving average
y2 = nan(T,1);       % initialize output vector with nan
for t=3:T-2          % since epsi is Tx1, t cannot start at 1 as we need epsi(t-2);
                     % t cannot end at T, as we end the loop with epsi(t+2);
    y2(t) = 1/5*(epsi(t-2)+epsi(t-1)+epsi(t)+epsi(t+1)+epsi(t+2));
end

%% Plots
f=figure('name','White Noise Plots');   % open new window for figure
subplot(1,2,1);                         % subplot(rows,columns, plotindex)
h1=plot(y1);                            % plot white noise
title('White Noise');                   % set title
subplot(1,2,2);                         % subplot(rows,columns, plotindex)
h2=plot(y2);                            % plot moving-average
title('5-point Moving Average');        % set title
```

Every simulation is different, model can thus generate an infinite set of realizations over the period $t = 1, ..., 200$. 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**:

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

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:

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

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:

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

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.
```
% =======================================================================
% ARPlots.m
% =======================================================================
% Willi Mutschler, October, 26 2022
% willi@mutschler.eu
% =======================================================================
clearvars; clc;close all;

%% Generate and plot autoregressive processes
phi=[-0.8 0.4 0.9 1.01];      % different values for the phi coefficient
sigma=1;                      % value for standard deviation of white noise,
    experiment with different values
T=200;                        % value for number of observations
Y=nan(T,4);                   % initiailize output vector with nan
Y(1,:)=zeros(1,4);            % set first period to zero

for j=1:4 % loop over coefficients
    for t=2:T % begin loop to compute AR(1) at t=2, as there is no y(0,j), i.e.
        you cannot index with 0
```

```
        Y(t,j)=phi(j)*Y(t-1,j)+randn()*sigma; % Simulate time seriers, randn simply generates one draw from N(0,1), we scale the standard deviation with sigma
    end
end

str_phi=["\phi=-0.8","\phi=0.4","\phi=0.9","\phi=1.01"]; % create array with titles of plots, note that MATLAB can handle (some) Latex expressions
% note the use of double quotes which creates a string array and it is easy to deal with strings of different length

figure('name','AR Plots'); % open new window for figure
for j=1:4 % loop over coefficients
    subplot(2,2,j);
    plot(Y(:,j));
    title(str_phi(j));
end


%% Generate and plot random walks
nRW=16;                  % number of Random Walks to generate
sigma=1;                 % value of standard error of white noise, experiment with different values
T=200;                   % number of observations
Y_RW=nan(T,nRW);         % initialize output vector with nan
Y_RW(1,:)=zeros(1,nRW); % set first period to zero
for j = 1:nRW
    for t=2:T
        Y_RW(t,j)=Y_RW(t-1,j)+randn()*sigma;
    end
end

figure('name','Random Walks: y_t = y_{t-1} + /varepsilon');
sgtitle('Random Walks')
for j=1:nRW
    subplot(4,4,j);
    plot(Y_RW(:,j));
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: $L^k y_t = L^{k-1}Ly_t = L^{k-1}y_{t-1} = ... = y_{t-k}$. Convenient use:

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

We can also work with lag-polynomials:

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

where we call p the lag order. So:

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

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 $(1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p) = 0$ (using a computer in most cases).

4 Solution to Properties of AR(1)

1. Let's first get a representation of the process that is useful to compute the moments. We can do this in different ways:

   - **Recursive substitution** (starting at some infinite time j):

   $$\begin{aligned}
   y_t &= \phi y_{t-1} + \varepsilon_t \\
   &= \phi\left(\phi y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t \\
   &= \phi^2(\phi y_{t-3} + \varepsilon_{t-2}) + \phi\varepsilon_{t-1} + \varepsilon_t \\
   &\vdots \\
   &= \phi^{j+1} y_{t-j+1} + \phi^j \varepsilon_{t-j} + ... + \phi^2\varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t
   \end{aligned}$$

   y_t is a linear function of an initial value $\phi^{j+1} y_{t-j+1}$ and historical values of the white noise process ε_t. If $|\phi| < 1$ and j becomes large, then $\phi^{j+1} y_{t-j+1} \to 0$, thus we get a so-called $MA(\infty)$ process:

   $$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ... = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

   - **With Lag Operators**: works only if $|\phi| < 1$ and $\{y_t\}$ is bounded, that is, there exists a finite number k such that $|y_t| < k$ for all t. Then

   $$\begin{aligned}
   (1-\phi L) y_t &= \varepsilon_t \\
   (1-\phi L)^{-1}(1-\phi L) y_t = y_t &= (1-\phi L)^{-1}\varepsilon_t
   \end{aligned}$$

   We can use the geometric rule: $(1-\phi L)^{-1} = \lim_{j\to\infty}(1 + \phi L + \phi^2 L^2 + ... + (\phi L)^j)$ Therefore:

   $$y_t = (1 + \phi L + \phi^2 L^2 + ... + (\phi L)^j)\varepsilon_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ... = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

   If we can express an AR process as a MA process, we call this process invertible. Let's now compute the moments from the MA(∞) representation and using the fact that ε_t is a white noise process:

   - **Unconditional Moments:**

   $$\begin{aligned}
   E[y_t] &= \mu = 0 \\
   Var[y_t] &= \gamma_0 = \frac{\sigma^2}{1-\phi^2} \\
   Cov[y_t, y_{t-k}] &= \gamma_k = \phi^k \gamma_0 \text{ for } k > 0
   \end{aligned}$$

   - **Conditional Moments:**
     - conditional on y_{t-1}:

     $$\begin{aligned}
     E[y_t|y_{t-1}] &= \phi y_{t-1} \\
     Var[y_t|y_{t-1}] &= \sigma^2 \\
     Cov[y_t, y_{t-j}|y_{t-k-1}] &= 0 \text{ for } k > 0
     \end{aligned}$$

     - conditional on y_{t-2}:

     $$\begin{aligned}
     E[y_t|y_{t-2}] &= \phi^2 y_{t-2} \\
     Var[y_t|y_{t-2}] &= (1+\phi^2)\sigma^2 \\
     Cov[y_t, y_{t-1}|y_{t-2}] &= \phi\sigma^2 \\
     Cov[y_t, y_{t-k}|y_{t-2}] &= 0 \text{ for } k > 1
     \end{aligned}$$

2.

```matlab
function r_k = ACFPlots(y,pmax,alph)
% ======================================================================
% r_k = ACFPlots(y,pmax,alph)
% ======================================================================
% Computes and plots the empirical autocorrelation function
% c_k = 1/T*\sum_{t=k+1}^T (y_t-\bar{y})(y_{t-k}-\bar{y})
% r_k = c_k/c0
% ----------------------------------------------------------------------
% INPUTS
%   - y    : Vector of data. [periods x 1]
%   - pmax : Maximum number of lags to plot. [scalar]
%   - alph : Significance level for asymptotic bands. [scalar, e.g. 0.05]
% ----------------------------------------------------------------------
% OUTPUTS
%   - rk   : Sample autocorrelation coefficient. [1 x maxlags]
% ======================================================================
% TEST CASE
% ARPlots
% r_k1 = ACFPlots(Y(:,1),8,0.05)
% r_k2 = ACFPlots(Y(:,2),8,0.05)
% r_k3 = ACFPlots(Y(:,3),8,0.05)
% r_k4 = ACFPlots(Y(:,4),8,0.05)
% you can compare to Matlab's econometrics toolbox:
% figure; autocorr(Y(:,1),'NumLags',8);
% ======================================================================
% Willi Mutschler, October 29, 2022
% willi@mutschler.eu
% ======================================================================

T=size(y,1);              % get number of periodes
y_demeaned = y-mean(y);   % put y in deviations from mean
r_k = nan(1,pmax);        % initialize output vector

% Compute variance
c0 = 1/T*(y_demeaned' * y_demeaned);
% Compute autocorrelations
for k=1:pmax
    c_k = 1/T * (y_demeaned(1+k:T,:)' * y_demeaned(1:T-k,:));
    r_k(1,k) = c_k/c0;
end

% Asymptotic bands
critval = norminv(1-alph/2);
ul = repmat(critval/sqrt(T),pmax,1);
ll = -1*ul;

% Barplots
figure('name','Autocorrelation');
bar(r_k);
hold on;
plot(1:pmax,ul,'color','black','linestyle','--');
plot(1:pmax,ll,'color','black','linestyle','--');
```

```
hold off;

% The following is just for pretty plots
acfbarplot = gca; % Get current axes handle
acfbarplot.Title.String = 'Sample autocorrelation coefficients';
acfbarplot.XAxis.Label.String = 'lags';
acfbarplot.XAxis.TickValues = 1:pmax;
acfbarplot.YAxis.Label.String = 'acf value';
acfbarplot.YAxis.Limits = [-1 1];
acfbarplot.XAxis.Limits = [0 pmax];

end % function end
```

5 Solution to Properties AR(1) With Time Trend

1.

$$
\begin{aligned}
y_t &= c + d \cdot t + \phi y_{t-1} + u_t \\
\Leftrightarrow y_t &= \frac{c + d \cdot t + u_t}{1 - \phi L} \\
\Leftrightarrow y_t &= \sum_{j=0}^{\infty} \phi^j (c + d(t-j) + u_{t-j}) = \frac{c}{1-\phi} + \frac{dt}{1-\phi} - d\sum_{j=0}^{\infty} \phi^j j + \sum_{j=0}^{\infty} \phi^j u_{t-j}
\end{aligned}
$$

as $|\phi| < 1$ the geometric series holds. For $\sum_{j=0}^{k} \phi^j j$ we also have a closed-form formula, which can be derived similar to the geometric series:

$$
\begin{aligned}
\text{Denote } S_k &= \sum_{j=0}^{k} \phi^j j = \sum_{j=1}^{k} \phi^j j = \phi^1 \cdot 1 + \phi^2 \cdot 2 + \phi^3 \cdot 3 + \cdots + \phi^k \cdot k, \\
\text{then } \phi S_k &= \phi^2 \cdot 1 + \phi^3 \cdot 2 + \phi^4 \cdot 3 + \cdots + \phi^{k+1} \cdot k \\
\text{and we get } (1-\phi) S_k &= \phi^1 \cdot 1 + \phi^2 \cdot 1 + \phi^3 \cdot 1 + \cdots + \phi^k \cdot 1 - \phi^{k+1} \cdot k \\
S_k &= \frac{1}{1-\phi} \sum_{j=1}^{k} \phi^j - \frac{\phi^{k+1}}{1-\phi} k.
\end{aligned}
$$

Looking at the limit of S_k for $k \to \infty$, we get $\lim_{k\to\infty} S_k = \frac{\phi}{(1-\phi)^2}$.
Therefore, y_t is given by

$$
y_t = \frac{c}{1-\phi} + \frac{dt}{1-\phi} - d\frac{\phi}{(1-\phi)^2} + \sum_{j=0}^{\infty} \phi^j u_{t-j}.
$$

Unconditional mean:

$$
\begin{aligned}
E[y_t] &= \frac{c}{1-\phi} + \frac{dt}{1-\phi} - d\frac{\phi}{(1-\phi)^2} + \underbrace{\sum_{j=0}^{\infty} \phi^j E[u_{t-j}]}_{=0, \text{ as } u_t \sim iid(0,\sigma^2)} \\
&= \frac{c}{1-\phi} + \frac{dt}{1-\phi} - d\frac{\phi}{(1-\phi)^2}
\end{aligned}
$$

Unconditional variance:

$$
\begin{aligned}
Var[y_t] &= E[(y_t - E[y_t])^2] \\
&= E[(\sum_{j=0}^{\infty} \phi^j E[u_{t-j}]) \cdot (\sum_{j=0}^{\infty} \phi^j E[u_{t-j}])] \\
&= E[(u_t + \phi^1 u_{t-1} + \phi^2 u_{t-2} + \dots)(u_t + \phi^1 u_{t-1} + \phi^2 u_{t-2} + \dots)] \\
&= E[u_t^2 + 2\phi^1 u_t u_{t-1} + 2\phi^2 u_t u_{t-2} + \dots + \phi^2 u_{t-1}^2 + 2\phi^3 u_{t-1} u_{t-2} + 2\phi^4 u_{t-1} u_{t-3} + \dots \\
&\quad + \phi^4 u_{t-2}^2 + 2\phi^5 u_{t-2} u_{t-3} + 2\phi^5 u_{t-2} u_{t-4} + \dots] \\
&\overset{iid}{=} E[u_t^2 + \phi^2 u_{t-1}^2 + \phi^4 u_{t-2}^2 + \dots]
\end{aligned}
$$

with $Var[u_t] = E[u_t^2] - E[u_t]^2 = E[u_t^2] = \sigma^2$ we get

$$
Var[y_t] = \sigma^2 (\phi^0 + \phi^2 + \phi^4 + \dots)] = \frac{\sigma^2}{1-\phi^2}.
$$

Autocovariance:

$$\begin{aligned}
\gamma(k) &= E[(y_t - E[y_t])(y_{t-k} - E[y_{t-k}])] \\
&= E[(\sum_{j=0}^{\infty} \phi^j u_{t-j})(\sum_{j=0}^{\infty} \phi^j u_{t-j-k})] \\
&= E[(u_t + \phi^1 u_{t-1} + \phi^2 u_{t-2} + \cdots + \phi^k u_{t-k} + \phi^{k+1} u_{t-k-1} + \phi^{k+2} u_{t-k-2} + \dots) \\
&\quad (u_{t-k} + \phi^1 u_{t-k-1} + \phi^2 u_{t-k-2} + \dots)] \\
&= E[u_t u_{t-k} + \phi^1 u_t u_{t-k-1} + \phi^2 u_t u_{t-k-3} + \dots \\
&\quad \phi^1 u_{t-1} u_{t-k} + \phi^2 u_{t-1} u_{t-k-1} + \phi^3 u_{t-1} u_{t-k-2} + \dots \\
&\quad \phi^2 u_{t-2} u_{t-k} + \phi^3 u_{t-2} u_{t-k-1} + \phi^4 u_{t-2} u_{t-k-2} + \dots \\
&\quad \vdots \\
&\quad \phi^k u_{t-k}^2 + 2\phi^{k+1} u_{t-k} u_{t-k-1} + 2\phi^{k+2} u_{t-k} u_{t-k-2} + \dots \\
&\quad \phi^{k+2} u_{t-k-1}^2 + 2\phi^{k+3} u_{t-k-1} u_{t-k-2} + 2\phi^{k+4} u_{t-k-1} u_{t-k-3} + \dots \\
&\quad \phi^{k+4} u_{t-k-2}^2 + 2\phi^{k+5} u_{t-k-2} u_{t-k-3} + 2\phi^{k+6} u_{t-k-2} u_{t-k-4} + \dots] \\
&\overset{iid}{=} E[\phi^k u_{t-k}^2 + \phi^{k+2} u_{t-k-1}^2 + \phi^{k+4} u_{t-k-2}^2 + \dots] \\
&= \phi^k \sigma^2 (\phi^0 + \phi^2 + \phi^4 + \dots) \\
&= \frac{\phi^k \sigma^2}{1 - \phi^2}
\end{aligned}$$

Autocorrelation: $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$. Therefore:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\frac{\phi^k \sigma^2}{1-\phi^2}}{\frac{\sigma^2}{1-\phi^2}} = \phi^k.$$

2. The expectation is time-dependent, hence it is not stationary. One could subtract the expectation, e.g. look at $y_t^s = y_t - E[y_t]$, where y_t^s is now covariance stationary. The unknown coefficient d must be estimated first though.