Exercise 1

Simon A. Broda

1. (a) Open the file `maunaloa.csv`; this is a famous data set used in machine learning. Make a time series plot.

   (b) Estimate a linear trend by regressing the `CO2` series on an intercept and a time trend. Hint: you can add a time trend to a dataframe `df` as follows.

   ```
   import statsmodels.tsa.api as tsa
   df = tsa.add_trend(df, trend="t")
   ```

   (c) Plot the data together with the estimated linear trend, and the residuals. An easy way to do this is

   ```
   import statsmodels.api as sm
   sm.graphics.plot_regress_exog(model, "trend");
   ```

   What do you notice?

   (d) Produce a forecast for Sept 1st, 2004 (one month after the sample ends), first manually using the fitted model

   $$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 t,$$

   then using Python.

   (e) Repeat Questions 1b through 1d, but using a quadratic trend.

   (f) Repeat Questions 1b through 1d, but using an exponential trend.

2. (a) Compute the 3rd order moving average of the `CO2` series for Feb 1st, 1964, both by hand and using Python. Hint: use a `rolling` object.

   (b) Estimate the trend with a 12 month moving average (12 months are necessary to cover a full cycle). Then plot the resulting trend estimate and the data together in a time series plot.

3. (a) Estimate a model with a linear trend and 12 monthly dummies (and no intercept) for the `CO2` series. Then, produce an (in-sample) forecast for August 1st, 2004, both by hand and using Python. Plot the data together with the estimated linear trend, and the residuals.

   (b) Same, but include an intercept. This will automatically remove one dummy.

Exercise 2

Simon A. Broda

1. (a) Open the file `simulations.xlsx`. The sheet "White Noise" simulates $T = 1000$ observations from a (Gaussian) white noise process; i.e., 1000 uncorrelated mean-zero normals. By repeatedly pressing `F9`, you can draw new random numbers. Describe your observations.

   (b) Similarly, the sheet "Random Walk" simulates $T = 1000$ observations from a (Gaussian) random walk. Describe your observations.

2. (a) Open the file `sp500.csv`. Add a column to the resulting dataframe with the name `log_price` containing the log prices, and a column `log_return` containing the continuously compounded returns. Make a time series plot for each, and a histogram of the returns. Describe your findings.

   (b) Compute the skewness and kurtosis of the returns and manually conduct a Jarque-Bera test. Then, use `statsmodels` to do the test.

   (c) Generate a correlogram (ACF and PACF) of the returns and interpret it.

   (d) Test whether the first 10 autocorrelations are jointly significant at the 5% level.

   (e) Generate a correlogram of the log prices and interpret it.

   Hint: Here are some useful functions.

   ```
   plt.hist
   sm.stats.stattools.robust_skewness
   sm.stats.stattools.robust_kurtosis
   sm.stats.stattools.jarque_bera
   scipy.stats.chi2.ppf
   sm.graphics.tsa.plot_acf
   sm.graphics.tsa.plot_pacf
   tsa.acf
   sm.stats.diagnostic.acorr_ljungbox
   ```

3. (a) Show that for the random walk $Y_t = Y_{t-1} + U_t$, where U_t is white noise and Y_0 some constant,

   $$Y_t = Y_0 + U_1 + U_2 + \cdots + U_t = Y_0 + \sum_{s=1}^{t} U_t.$$

   (b) Building on the result from the previous question, show that

   $$\mathbb{E}[Y_t] = Y_0, \quad \text{and}$$
   $$\text{var}(Y_t) = \sigma^2 t.$$

Exercise 3

Simon A. Broda

1. (a) Open the file `simulations.xlsx`. The sheet "AR(1)" simulates $T = 1000$ observations from an AR(1) process. Play around with α and $-1 < \phi_1 < 1$ and describe your observations.

   (b) Also try setting $\phi_1 = 1$ and describe the effect of α.

   (c) The file `simulated_data.csv` contains three series simulated using the same spreadsheet, `simulation.xlsx`, one each for an AR(1), an MA(1), and an ARMA(1, 1) process. The AR and ARMA processes use $\phi_1 = 0.7$, and the MA and ARMA processes use $\theta_1 = 0.7$. Plot the sample ACF and PACF for both, and describe your observations.

2. (a) Use the Box-Jenkins approach to model year-on-year real GDP growth. The (quarterly) GDP data can be found in the file `realgdpch.csv`. You will need to transform them into year-on-year growth rates first, by doing

```
df["YoY_GROWTH"] = np.log(df["REAL_GDP"]) \
                 - np.log(df["REAL_GDP"]).shift(4)
```

   (b) Produce forecasts for 2022Q3 and 2022Q4, both manually and using Python. Do this for

      i. an MA(3) model, and

      ii. an AR(1) model.

3. (a) Obtain the mean and variance of a random walk with drift.

   (b) Show that the random walk with drift is integrated of order 1.

   (c) Derive the expression for the variance of a stationary AR(1) given in the slides.

   (d) **Optional**: Find the mean, variance, and ACF of an MA(1).

   (e) **Optional**: Find the ACF of a stationary AR(1).

Exercise 4

Simon A. Broda

1. Open the file `simulations.xlsx`. Use the sheets "AR(1)" (with ϕ_1 set to 1) to simulate a random walk with drift, and the sheet "Linear Trend" to simulate a trend-stationary process. Play with the parameters and describe your observations.

2. (a) The file `tbill.csv` contains monthly data for the 3-month T-Bill rate. Plot them, study the correlogram, and conduct a unit root test.
   (b) Model the first difference of the T-Bill rate as an ARMA process, hence modelling the T-Bill rate as an ARIMA process.
   (c) Forecast the T-Bill rate for 2022M11 and 2022M12 based on the model you found in the previous question.

3. (a) Show that for both

$$
\begin{aligned}
Y_{1,t} &= \delta t + U_{1,t} \quad \text{and} \\
Y_{2,t} &= \delta + Y_{2,t-1} + U_{2,t}.
\end{aligned}
$$

   we have $\mathbb{E}[\Delta Y_{i,t}] = 0$.
   (b) Derive the ADF regression for an AR(2) process.

Exercise 5

Simon A. Broda

1. (a) Load the data in the file `sp500.csv` and construct the log returns (in percent). Make a time series plot and a histogram of the returns. Then, produce a correlogram of the squared residuals, and test the first 5 autocorrelations of the squared returns for joint significance using the Ljung-Box Q-test. Interpret your results.

   (b) Perform an ARCH-LM test by regressing the returns on an intercept, and feeding the residuals into `sm.stats.diagnostic.het_arch`.

   (c) Compute the historical volatility and plot it, by using a rolling object.

   (d) Compute the EWMA volatility and plot it, by using a ewm object. Also, try and recreate the plot with the EWMA volatilities for different values of λ shown in the slides.

   (e) Find a suitable (G)ARCH or TARCH model. Start with a GARCH(1, 1) or an ARCH(6) model, and determine whether it needs to be adjusted. You may need to install the `arch` package (docs). This can be done by putting `!pip install arch` into a Jupyter cell and executing it.

   (f) Make a plot of the volatility estimates that your model generates. Then, compute the unconditional (or average) volatility, and plot the NIC.

   (g) Forecast the volatility for $T + 1$.

2. (a) Show that

$$\widehat{\sigma}^2_{t+1,EWMA} = \lambda\widehat{\sigma}^2_{t,EWMA} + (1-\lambda)r_t^2, \qquad 0 < \lambda < 1.$$

Exercise 6

Simon A. Broda

1. (a) Load the data in the file `sp500.csv` and construct the log returns (in percent). Use a `rolling` object with a window of $m = 250$ days to compute and plot the historical 1% VaR.

   (b) Determine the Normal VaR for the entire sample, both using Python and manually. Also compute (and plot) the normal VaR with the mean and volatility estimated from a rolling window of length $m = 250$.

   (c) Determine the VaR based on a GARCH(1, 1) model with normal innovations.

   (d) Same, but with standardized t innovations.

   (e) Predict the VaR (manually and using Python) for 10/28/2022 based on the GARCH model with t innovations.

   (f) Make a plot with your VaR estimates overlaid on the negative log returns.

   (g) Test your VaR forecasts for correct unconditional coverage, independence, and correct conditional coverage.

Exercise 7

Simon A. Broda

1. Until 1971, as part of the Bretton-Woods system of fixed exchange rates, the US dollar was convertible to gold, i.e., it was possible for foreign central banks to redeem US dollars for gold at a fixed rate of 35$ per troy ounce, so that the price of gold was fixed. In 1971, US president Nixon unilaterally cancelled the direct convertibility, ultimately ending the Bretton-Woods agreement. Gold became a floating asset, and its price increased sharply; in other words, the US$ was massively devalued. In this exercise, we will analyze the hypothesis that the increasing price (in US$) of oil is not a consequence of an increased demand for (or a reduced supply of) oil, but rather of a continued devaluation of the US$. We have at our disposal monthly data from April 1968 to January 2017 (586 observations) on the following variables:

   - GOLD, the spot price of one troy ounce of gold in US$;
   - OIL, the spot price of one barrel of WTI crude oil in US$.

   (a) Assuming that GOLD is integrated of order one, explain why the hypothesis that the relative price of oil (in troy ounces of gold per barrel) is stationary implies cointegration between log(OIL) and log(GOLD).

   (b) Using the file `oil_gold_2017.csv`, analyze whether this cointegrating relationship can be found in the data, based on the Engle-Granger procedure. The following steps are required:

      i. Transform the data into logs, and plot the two resulting series together. What do you notice?

      ii. Perform an ADF test for both series. Make sure to specify the deterministic regressors (constant and/or trend) correctly. What do you conclude?

      iii. Estimate the long-run relationship (cointegrating relationship)

      $$\text{loil}_t = \beta_1 + \beta_2 \text{lgold}_t + U_t.$$

      State the cointegrating vector, and make a plot of the residuals.

      iv. Perform the Engle-Granger test, i.e., apply an ADF test to the residuals $\hat{u}_t$. What do you conclude?

      v. Estimate a vector error correction model. Write the two estimated equations out.

2. Consider the model

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + U_{1,t} \\ X_t &= X_{t-1} + U_{2,t} \end{aligned}$$

where $\beta_2 \neq 0$, $U_{1,t}, U_{2,t} \overset{\text{iid}}{\sim} (0, \sigma^2)$ independently of each other.

(a) Is X_t stationary?

(b) Is Y_t stationary?

(c) Are X_t and Y_t cointegrated? If yes, what is the cointegrating vector?

(d) Derive the bivariate VECM for Y_t and X_t.