Module 9.3: Time Series Analysis with Python
Fall Term 2024

Week 1:

Introduction; Descriptive Time Series Analysis

HSLU Lucerne University of Applied Sciences and Arts

Outline in Weeks

1. Introduction; Descriptive Modeling
2. Returns; Autocorrelation; Stationarity
3. ARMA Models
4. Unit Roots; ARIMA Models
5. Volatility Modeling
6. Value at Risk
7. Cointegration

Outline

General Information

- Lectures will be a mix of theory and practice.
- Slides and additional materials are available on Ilias.
- 90 min. written exam during exam phase, closed book. Details will be communicated later.
- A mock exam will be made available.

Book

- Course is not explicitly based on any book.
- If you prefer to have a book, then I recommend Brooks (2019)[1]. A reading list follows on the next slide.
- I will also make selected problems and solutions available.

[1] Brooks, C. (2019). Introductory Econometrics for Finance (4th ed.). Cambridge: Cambridge University Press.

Reading List: Brooks (2019)

- **Pre** As a refresher, Sections 1.1–1.6 (mathematical foundations), 2.1–2.7 (statistics and distribution theory).
- **Week 1** (not covered in book)
- **Week 2** Sections 6.1 and 6.2
- **Week 3** Sections 6.3–6.10
- **Week 4** Section 8.1; Section 5.5
- **Week 5** Sections 9.1–9.16; 9.17
- **Week 6** (not covered in book)
- **Week 7** Sections 8.3 – 8.11
- **Week 7** Sections 11.1–11.7; Section 14.2

Outline

What is Time Series Analysis

- In Module 9.2, you learned about the classical linear regression model (*CLRM*).
- It is used to estimate linear relationships of the form

$$Y_i = \beta_0 + \beta_1 x_i + U_i, \qquad (\dagger)$$

  possibly with more than one regressor.
- Typically, 5 assumption are made about the error term: zero mean, constant variance (homoskedasticity), lack of autocorrelation, no correlation with the regressors (orthogonality), normality.
- These are often justifiable for *cross-sectional* data, where each observation i corresponds to a different entity (e.g., a firm, a country, etc.).

What is Time Series Analysis

- In many areas of financial econometrics (risk models, asset pricing, ...), one deals with *time series data* instead; here, every observation corresponds to a different *time period*. Examples:
    - the price of IBM stock on each trading day since Jan 2nd, 2004;
    - monthly inflation in the EUR area since Jan 2002;
    - US GDP growth in every quarter since 1986Q1, etc.
- As seen above, time series may have different *frequencies* (daily, monthly, quarterly, etc.).
- We will only cover *regular* time series: observations occur at equally spaced time points (e.g., daily closing prices for stocks).

What is Time Series Analysis

- To highlight the fact that we are dealing with time series, we use a subscript t instead of i; thus, a regression model such as † would be written

$$Y_t = \beta_0 + \beta_1 x_t + U_t \qquad (\ddagger)$$

  if $\{Y_t\}$ and $\{x_t\}$ are time series.
- Regression ‡ is unlikely to satisfy the CLRM assumtions; time series usually exhibit *autocorrelation*, and often changes in standard deviation (or in "volatility", for stock returns).
- Time series analysis is the study of methods to deal with these salient features.
- The broader goal (as usual in econometrics) is to empirically *verify* economic theories (e.g., the CAPM).
- Another important aspect is *forecasting* (e.g, GDP forecasts, inflation forecasts, Value at Risk forecasts, etc.)

What is Time Series Analysis

- For most of the course, we will consider *univariate* time series analysis.
- This means that instead of a regression like

$$Y_t = \beta_0 + \beta_1 x_t + U_t$$

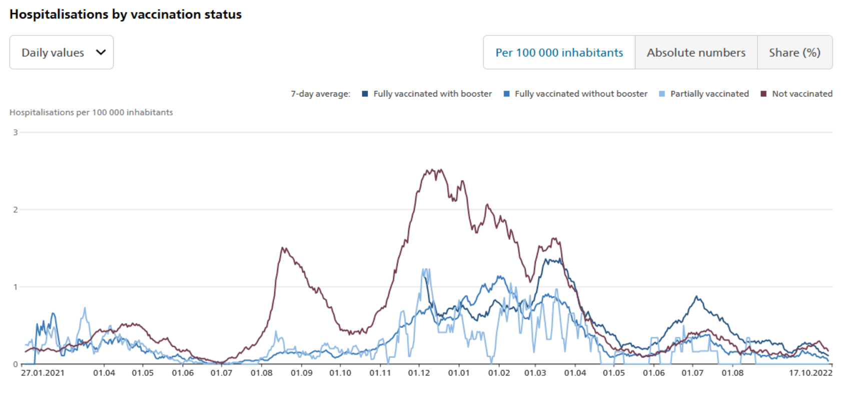
  above, we only have *one* time series $\{Y_t\}$.
- The goal is to describe the (dynamic) behavior of Y_t, e.g., for forecasting.
- We'll start with a purely *descriptive* approach today. Starting next week, we'll move on to actual dynamic models.

Outline

Bitcoin prices

Source: coinmarketcap.com

CoViD19 Hospitalizations

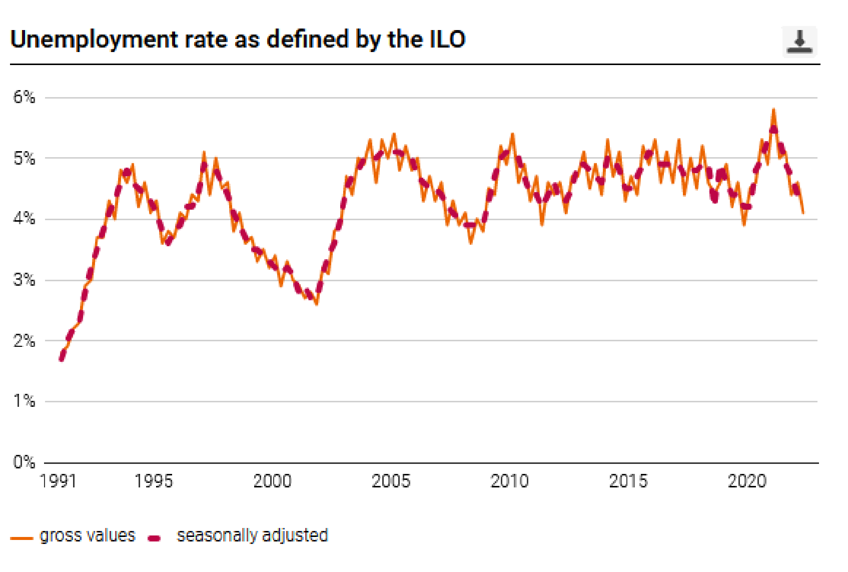


Source: covid19.admin.ch

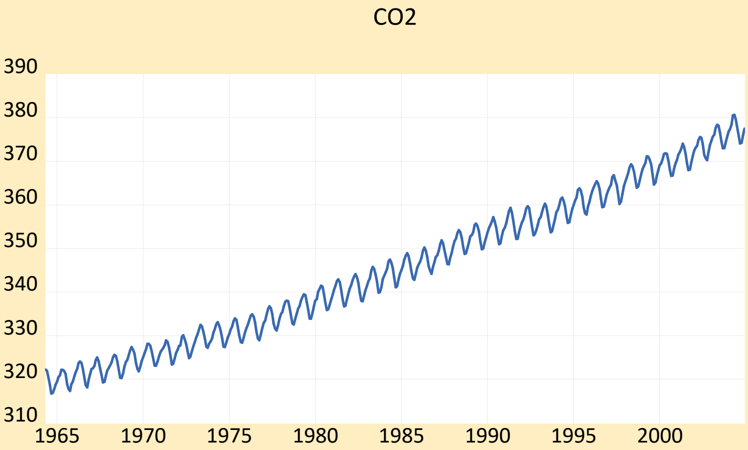
Tesla Stock

Source: Yahoo Finance

Unemployment in Switzerland



Atmospheric CO2 on Mauna Loa, Hawaii



Outline

Time Series Plots

- The above plots were all examples of *time series plots*: plotting the data against time itself.
- This is usually the first thing to do when looking at a new data set.
- We'll see later how to make these plots.

Decomposing a Time Series

- The main tool of descriptive time series analysis it to decompose it into a *trend*, a *seasonal* component, and a *residual* component, according to the additive model

$$Y_t = F_t + S_t + U_t,$$

  where the trend component F_t models long-term movements, the seasonal component S_t measures systematic seasonal patterns, and the residual component U_t contains anything that cannot be explained by the other two[2].
- The Mauna Loa data make the trend and seasonal component very obvious.

[2] Sometimes economic time series also contain a cyclical component stemming from the business cycle, but we will ignore this here.

Outline

Estimating a Linear Trend by OLS

- One way to estimate a *linear trend* is to just regress the data on an intercept and time itself, i.e.,

$$Y_t = \beta_0 + \beta_1 t + U_t.$$

- The estimated trend is then

$$\widehat{F}_t = \widehat{\beta_0} + \widehat{\beta_1} t.$$

Estimating a Quadratic Trend by OLS

- As seen in the exercises, it is also possible to have a nonlinear trend. One example is a *quadratic* trend. This can be estimated via the regression

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + U_t.$$

- The estimated trend is then

$$\widehat{F}_t = \widehat{\beta_0} + \widehat{\beta_1} t + \widehat{\beta_2} t^2.$$

Estimating an Exponential Trend by OLS

- Another possibility is to use an *exponential* trend. The model is then

$$F_t = \beta_0 \cdot \beta_1^t.$$

- To estimate this by OLS, one takes logs:

$$\log(F_t) = \log(\beta_0) + \log(\beta_1) \cdot t =: c + b \cdot t.$$

- Adding an error term, this exponential trend can be estimated via the regression

$$\log(Y_t) = c + b \cdot t + U_t.$$

- The resulting trend function is

$$F_t = \widehat{\beta_0} \cdot \widehat{\beta_1}^t, \quad \text{where } \widehat{\beta}_0 = \exp(\hat{c}), \widehat{\beta}_1 = \exp(\hat{b}).$$

Interpreting an Exponential Trend

- If the trend is
$$F_t = \beta_0 \cdot \beta_1^t,$$
then
$$\frac{F_t}{F_{t-1}} = \frac{\beta_0 \cdot \beta_1^t}{\beta_0 \cdot \beta_1^{t-1}} = \beta_1;$$
i.e., Y_t grows by $100 \cdot (\beta_1 - 1)\%$ per period, on average.
- Example ($\beta_0 = 1, \beta_1 = 1.05$):
$$F_t = 1.05^t,$$
so Y_t grows by 5% a year, on average (cf. compounding interest).

Estimating the Trend via Moving Averages

- Another approach, which has the advantage of adapting to the data automatically, rather than pre-specifying a functional form (linear, quadratic, exponential), is to estimate the trend via a *moving average.*
- E.g., for a third-order moving average ($k = 3$),

$$\widehat{F}_t = (Y_{t-1} + Y_t + Y_{t+1})/3.$$

- Choice of k: the higher, the smoother. If seasonality is present, k should cover at least a full cycle.
- *Downside*: $(k + 1)/2$ values at the end points cannot be computed. Thus also not useful for forecasting.
- Note: for a moving average of even order, one averages $k + 1$ data points, but the endpoints get half the weight. E.g., with $k = 4$,

$$\widehat{F}_t = \left(\frac{1}{2}Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + \frac{1}{2}Y_{t+2}\right)/4.$$

Outline

Dummy Variables

- Simple way to address the seasonality S_t: *seasonal dummies*, which take the value one in one season, and zero in all others.
- Example (see next page): say we have quarterly data. Then we would use four dummies, defined, for $j \in \{1, \ldots, 4\}$, as

$$d_{jt} = \begin{cases} 1, \text{ if observation } t \text{ is in season } j, \\ 0, \text{ otherwise.} \end{cases}$$

- Effectively, every season gets its own intercept.
- Careful: if a full set of dummies is included, then the intercept must be left off, otherwise the regressors are perfectly collinear; this is the *dummy variable trap*.
- Alternatively, keep the intercept, but remove one of the dummies. That season then becomes the baseline, and the other dummies measure the average difference from the baseline, per season.

Example

t	Date	d_1	d_2	d_3	d_4
1	2021Q1	1	0	0	0
2	2021Q2	0	1	0	0
3	2021Q3	0	0	1	0
4	2021Q4	0	0	0	1
5	2022Q1	1	0	0	0
6	2022Q2	0	1	0	0
7	2022Q3	0	0	1	0
8	2022Q4	0	0	0	1
$\vdots$					

Example continued

- If we include a linear trend, then the model becomes

$$\begin{aligned} Y_t &= F_t + S_t + U_t \\ &= \beta_1 \cdot t + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 d_{3,t} + \alpha_4 d_{4,t} + U_t, \end{aligned}$$

  which can be estimated by OLS.
- If we want to produce a forecast for Y_6, which is in Season 2, then

$$\widehat{Y_6} = \widehat{\beta_1} \cdot 6 + \widehat{\alpha_2}.$$

Example continued

- Alternatively, include an intercept and drop one dummy:

$$Y_t = \beta_0 + \beta_1 \cdot t + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 d_{3,t} + U_t.$$

- This makes Season 4 our baseline; the other seasons are measured in deviation from this baseline.
- The forecast for an observation in Season 4 is thus simply

$$\widehat{Y_4} = \widehat{\beta_0} + \widehat{\beta_1} \cdot 4$$

- The other seasons are measured in deviation from the baseline; e.g., α_2 is the average difference between Seasons 4 and 2:

$$\widehat{Y_6} = \widehat{\beta_0} + \widehat{\beta_1} \cdot 6 + \widehat{\alpha_2}.$$

Outline

Learning Goals

Students

- Know the difference between cross-sectional and time series data,
- know what a regular time series is, and what its frequency is,
- are able to decompose a time series into trend, seasonality, and the residual component using Python,
- and are able to produce time series plots in Python.

Homework

- Freshen up your statistics knowledge, if needed.
- Exercise 1.