Chapter 1
Introduction

1.1 Overview

This course is designed to provide an overview of time series analysis using the Python language. We will introduce some of the common time series analysis techniques as well as how to implement and understand them in Python. In addition we will cover some of the common manipulation tasks related to dates and times and how we can create visualisations to aid our analysis.

1.2 Course Materials

Items appearing in this material are sometimes given a special appearance to set them apart from regular text. Here's how they look:

```
>>> This is a section of code              # This is a comment
```

A warning, typically describing non-intuitive aspects of the Python language

A tip: additional features of Python or "shortcuts" based on user experience

Exercises to be performed during (or after) the training course

1.3 Course Script and Exercise Answers

A great deal of code will be executed within Python by the consultant during the delivery of this training. This includes the answers to each exercise, as well as other code written to answer questions that arise. Following the course, each attendee will be sent a script containing all the code that was executed, annotated with additional comments.

1.4 What is a Time Series?

A time series is a sequence of values that are generated by a statistical process that is dependent on a time dimension. The analysis and modelling of time series is different from other statistical learning techniques in that the time dimension is not explicitly modelled. Weekdays, weekends, summer months, leap years, etc. are not variables we use. Instead the time dimension is captured implicitly through the values of the time series itself. This way we try to model recurring patterns (seasonality) and consistent changes (trend).

In this course we will cover the basics of time series analysis and modelling. More advanced concepts such as state space models or neural networks are beyond the scope of this course.

1.5 Packages

In the python ecosystem of packages for data analysis, **statsmodels** is the go-to package for anything related to statistical modelling. This includes topics such as linear regression, statistical tests and time series analysis. With regard to the latter, **statsmodels** includes implementations of the most common algorithms which we will cover in this course.

Besides **statsmodels** we will also cover the time series capabilities of the **pandas** package. This package has become the de facto standard when it comes to data analysis in python. It includes functionality for reading, transforming and analysing data. The author of **pandas** worked in the financial industry whilst developing the package, it therefore also includes functionality to analyse and easily visualise time series.

There are other python packages available for time series analysis and modelling (e.g. **pyflux**) but these mostly cover more advanced topics.

1.6 Further Reading

There are a variety of resources available to get you started with time series analysis. We list some of the more useful ones here:

- Hyndman, Rob J., and George Athanasopoulos (2018). *Forecasting: principles and practice*. OTexts.
- Brockwell, Peter J., and Davis, Richard A (2016). *Introduction To Time Series And Forecasting*. 3rd edition. New York: Springer.
- Chatfield, Chris (2008). *The Analysis Of Time Series An Introduction*. 5th edition. Chapman & Hall/CRC.