Lecture Note 1 - Introduction

This course provides the basic tools needed to analyze data that is observed in a sequential ordering. Standard inferential techniques no longer work in this setup without further qualification since the data and the innovations no longer can be viewed as independent. This requires a refined limit theory to allow for statements about asymptotic properties of estimators and test statistics. One possible conceptual solution to the dependence problem lies in the formulation of generating mechanisms, that allow to decompose the observed dependent data into independent innovations. These mechanisms are usually referred to as time series models. While other less parametric approaches exist we will mainly focus on these parametric formulations.

To focus ideas we consider a few macroeconomic time series such as GNP or aggregate consumption. Typical features displayed by these time series include trending behavior, seasonality and cyclical components around trends. It is clear that these concepts are vague, and in fact can not be uniquely defined. They serve as a conceptual guideline to building time series models. In particular modelling trending behavior is a key issue. Without the notion of some type of stability, or, in more technical language, stationarity, of the data, it would be impossible to make inferential statements.

1. Stochastic Processes and Stationarity

To formalize the discussion of these concepts we start by defining what we mean by a time series and by introducing concepts of stationarity. The mathematical theory behind time series analysis is based on the notion of there being an abstract probability space \((\Omega, \mathcal{F}, P)\). Here \(\Omega\) is the sample space, \(\mathcal{F}\) is a sigma algebra defined on the sample space and \(P\) is a probability measure on \(\Omega\). For most of the arguments in the course, reference to the underlying probability space is not necessary and a discussion of its properties is omitted for this reason. We introduce an index \(T\) which has an ordering relation defined on it. So if \(t_1, t_2 \in T\) then \(t_1 \leq t_2\) or \(t_1 > t_2\). Usually \(T = \mathbb{R}\) or \(T = \mathbb{Z}\). We give a definition for a stochastic process first.

**Definition 1.1 (Stochastic Process).** A stochastic process is a family of random variables \(\{X(t, \omega), t \in T, \omega \in \Omega\}\) defined on a probability space \((\Omega, \mathcal{F}, P)\).

In particular for a fixed \(\omega\), \(X(., \omega)\) is a function from \(T\) into \(\mathbb{R}\). This is called a realization of the stochastic process. On the other hand, for \(t\) fixed, \(X(t, .)\) is a function from \(\Omega\) into \(\mathbb{R}\). The definition contains continuous as well as discrete time processes. We call a process a time series, if the index \(T\) is discrete as is the case for \(\mathbb{Z}\). A time series can be generated from a stochastic process by looking at a grid of points in \(T\). We define \(\mathbb{R}^\infty = \bigotimes_{-\infty}^{\infty} \mathbb{R}\) to be the infinite dimensional coordinate space and represent the sequence \(\{X(., \omega)\}_{t=-\infty}^{\infty}\) as a point in \(\mathbb{R}^\infty\).

**Definition 1.2 (Random Sequence).** A random sequence is defined as a mapping \(h : \Omega \rightarrow \mathbb{R}^\infty\) such that

\[
h(\omega) = (..., X_{-1}(\omega), X_0(\omega), X_{-1}(\omega), ...)
\]

where \(X_t : \bigotimes_{-\infty}^{\infty} \mathbb{R} \rightarrow \mathbb{R}\) are the coordinate functions. \(\{X_t\}_{-\infty}^{\infty}\) is called a random sequence.
In the same way as before we call \( h(\omega) \) a realization of the random sequence for a given \( \omega \). In reality we only observe one realization of \( h(\omega) \) and we only see a finite subset of the \( X_t \), i.e. a sequence \( \{X_t\}_{t=1}^{\infty} \). We can therefore only hope to make reasonable inference about the probability distribution of \( X_t \) if this distribution does not change in an unknown way over time. We introduce the following concepts of strong and weak stationarity. Before that, we need to make the notion of distribution of \( X_t \) more precise.

**Definition 1.3 (Finite Dimensional Distributions).** Let \( T \) be the set of all vectors

\[
\{t = (t_1, \ldots, t_n), t_1 < \ldots < t_n, n = 1, 2, \ldots\}.
\]

The finite dimensional distribution functions of \( \{X_t, t \in T\} \) are the functions \( \{F_t(\cdot), t \in T\} \) defined by

\[
F_t(x) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) \text{ where } x = (x_1, \ldots, x_n) \in \mathbb{R}^n.
\]

Using this definition we can now state what we mean by a strictly stationary process.

**Definition 1.4 (Strict Stationarity).** A random sequence \( \{X_t\}_{t=1}^{\infty} \) is strictly stationary if the finite dimensional distributions are translation invariant, i.e.

\[
(X_{t_1+h}, \ldots, X_{t_n+h}) \overset{d}{=} (X_{t_1}, \ldots, X_{t_n}) \forall h, n, t.
\]

**Remark 1.** Strict stationarity is a stronger concept than the identical distribution assumption since it requires that all marginals have to be identical.

**Example 1.5.** Let \( \{X_t\}_{t=-\infty}^{\infty} \) such that \( Z_n = (X_{2n-1}, X_{2n}) \) where \( Z_n \sim iidN(0, \Sigma) \) with

\[
\Sigma = \begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}.
\]

Then \( X_t = N(0,1) \forall t \) but \( (X_1, X_2) \not\overset{d}{=} (X_2, X_3) \) since \( (X_2, X_3) = N(0, I) \).

Joint finite dimensional distributions are difficult to work with. It is therefore useful to introduce a simpler concept of stationarity which only characterizes the first two moments of the process. A very important concept in time series analysis is the autocovariance function. The autocovariance function is a measure of dependence between different elements of the sequence \( \{X_t\}_{t=-\infty}^{\infty} \).

**Definition 1.6 (Autocovariance Function).** Let \( \{X_t\}_{t=-\infty}^{\infty} \) be such that \( EX_t^2 < \infty \) for all \( t \). Then the autocovariance function is given by

\[
\gamma_{XX}(t, s) = E(X_t - EX_t)(X_s - EX_s).
\]

The covariance function has certain properties. In particular \( \gamma_{XX}(t, t) \geq 0 \) and \( |\gamma_{XX}(t, s)| \leq \sqrt{\gamma_{XX}(t, t)\gamma_{XX}(s, s)} \) by the Cauchy-Schwarz inequality. If the stochastic process is stationary in the following weak sense then we have an additional property.

**Definition 1.7 (Stationarity).** The time series \( \{X_t\}_{t=-\infty}^{\infty} \) is said to be weakly stationary (covariance stationary) if
1. $EX_t^2 < \infty$ for all $t$

2. $EX_t = c$ for all $t$ where $c \in \mathbb{R}$

3. $\gamma_{XX}(t, s) = \gamma_{XX}(t + r, s + r)$ for all $r, t$ and $s$.

We can choose $r = -t$ such that $\gamma_{XX}(t, s) = \gamma_{XX}(0, s - t)$. It is therefore common to define $\gamma_{XX}(h) = \text{Cov}(X_{t+h}, X_t)$ if $X_t$ is covariance stationary. It now follows that the covariance function of a stationary process is even or $\gamma_{XX}(h) = \gamma_{XX}(-h)$.

**Definition 1.8.** A real valued function $f(h) : \mathbb{Z} \rightarrow \mathbb{R}$ is non-negative definite if for any $n$ and all vectors $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $t = (t_1, \ldots, t_n) \in \mathbb{Z}^n$ it follows

$$
\sum \sum a_i a_j f(t_i - t_j) \geq 0
$$

It can be shown that $\gamma_{XX}(h)$ is non-negative definite. We discuss several relationships between the two stationarity concepts next.

**Remark 2.** If $\{X_t\}_{t=-\infty}^{\infty}$ is such that $EX_t^2 < \infty$ for all $t$ then strict stationarity implies covariance stationarity.

**Remark 3.** In general covariance stationarity does not imply strict stationarity. The exception are Gaussian processes. A process $\{X_t\}_{t=-\infty}^{\infty}$ is Gaussian if all joint distributions are multivariate Gaussian. Then if $\{X_t\}_{t=-\infty}^{\infty}$ is weakly stationary and Gaussian, it is also strictly stationary since by weak stationarity $(X_{t_1+h}, \ldots, X_{t_n+h})$ and $(X_{t_1}, \ldots, X_{t_n})$ have the same mean and covariance matrix. By Gaussianity this implies that they have the same distribution.

### 2. Modeling Trends

The concepts of stationarity typically do not apply to raw data series observed in economic contexts. Economic data usually follows trends. Various approaches to handle trending behavior have been proposed in the literature. One typically distinguishes deterministic and stochastic trends. A simple model of a deterministic trend is the following polynomial trend model. Let $\{Y_t\}_{t=1}^{T}$ be the observed time series generated by

$$
Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_p t^p + U_t
$$

where $u_t$ is a stationary process. An important special case is the linear time trend model with $p = 1$. If we define $X_t = (1, t, t^2, \ldots, t^p)$ then we can write (2.1) as a standard regression equation $Y_t = X_t \beta + U_t$. In general we know that the OLS estimator is inefficient if $\text{Cov}(u_t, u_s) \neq 0$ for $t \neq s$. A famous result by Grenander and Rosenblatt (1957) shows that if $X_t = (1, t, t^2, \ldots, t^p)$ then asymptotically $\text{OLS} = \text{GLS}$. This means that $Y_t$ can be efficiently detrended by OLS. More explicitly, let $X = (X_1, \ldots, X_T)'$ and $Y = (Y_1, ..., Y_n)$. Then $\hat{Y} = (I - X(X'X)^{-1}X')Y$. The transformation removes the trend in $Y_t$ but does not generate stationarity.

An alternative detrending method is to take first differences. We look at

$$
\Delta Y_t = Y_t - Y_{t-1} = \beta_1 + U_t - U_{t-1}.
$$

It is left as an exercise to show that this transformation leads to a stationary process.

Stochastic trends are an alternative representation of trending variables. An important case is the random walk model.
Example 2.1 (Random Walk). Let \( \{U_t\}_{t=1}^T \) be a sequence of independently and identically distributed random variables with \( EU_t = 0 \) and \( EU_t^2 < \infty \). A sequence \( \{S_t\}_{t=1}^T \) with \( S_0 = 0 \) and

\[
S_t = \sum_{i=1}^t U_i
\]

is called a random walk.

We see that \( ES_t = 0 \) but \( Var(S_t) = t\sigma_u^2 \) such that \( S_t \) is nonstationary. \( S_t \) is said to follow a stochastic trend since it is \( O_p(\sqrt{T}) \). It is immediate that the transformation \( \Delta S_t = U_t \) leads to a stationary process.

The previous model of a random walk can easily be generalized in several ways. The one that is most relevant to our discussion is the random walk plus drift model.

Example 2.2 (Random Walk plus Drift). Let \( \{U_t\}_{t=1}^T \) be a sequence of independently and identically distributed random variables with \( EU_t = 0 \) and \( EU_t^2 < \infty \). A sequence \( \{S_t\}_{t=1}^T \) with \( S_0 = 0 \) and

\[
\Delta S_t = \mu + U_t
\]

such that

\[
S_t = \mu t + \sum_{i=1}^t U_i
\]

is called a random walk with drift since now \( ES_t = \mu t \).

We note that \( S_t = O_p(t) \). This shows that the stochastic trend is dominated by the deterministic trend. By the same arguments as before it can be seen, that taking differences leads to a stationary process while filtering out the time trend by projecting it on \( X_t \) does not.

We will discuss the consequences for inference of the different forms of trends in more detail at a later stage. For expositional purposes we consider however the following example. Let \( x_t = \mu t + \sum_{i=1}^t u_i \) with \( x_0 = 0 \) and assume that we have the following regression equation

\[
y_t = \beta x_t + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim iid(0, \sigma^2) \quad \text{and} \quad \varepsilon_t, u_i \text{ independent for all } t, i
\]

Then the OLS estimator for \( \beta \) converges at the rate \( T^{3/2} \). This follows from

\[
\frac{1}{T^3} \sum_{t=1}^T x_t^2 = \frac{1}{T^3} \sum_{t=1}^T (\mu t^2 + 2\mu t \sum_{i=1}^t u_i + (\sum_{i=1}^t u_i)^2) \to \mu^2 / 3.
\]

(2.2)

and

\[
T^{3/2}(\hat{\beta} - \beta) = \frac{1}{T^{3/2}} \sum_{t=1}^T \frac{\mu t \varepsilon_t}{x_t^2} = \frac{1}{T^{3/2}} \sum_{t=1}^T \frac{\mu t \varepsilon_t}{x_t^2} + o_p(1) = O_p(1).
\]

(2.3)

It is illustrative to consider what would happen if we estimate the model in first difference form. This transformation is often advocated to reduce the data to a stationary form. We consider the following regression

\[
\Delta y_t = \beta \Delta x_t + \varepsilon_t - \varepsilon_{t-1}
\]

\[
= \beta (\mu + u_t) + \varepsilon_t - \varepsilon_{t-1}
\]
Then, the OLS estimator on the differenced data is
\[ \tilde{\beta} = \beta + \sum (\mu + u_t)(\varepsilon_t - \varepsilon_{t-1}) / \sum (u_t + \mu)^2. \]

Now by a law of large numbers for iid random variables \( \frac{1}{T} \sum u_t^2 \to \sigma_u^2 \) such that \( \frac{1}{T} \sum (u_t + \mu)^2 \to \sigma_u^2 + \mu^2 \). Also \( E_{u_t}(\varepsilon_t - \varepsilon_{t-1}) = 0 \) and

\[ Var \left( \frac{1}{T^{1/2}} \sum (\mu + u_t)(\varepsilon_t - \varepsilon_{t-1}) \right) = 2\sigma_u^2 \sigma_\varepsilon^2 \]

which shows that \( (\tilde{\beta} - \beta) = O_p(T^{-1/2}) \). This example shows that taking differences to enforce stationarity is not always a good strategy. Differentiating removes the level information from the data. In this sense the differenced regression is an order of magnitude less informative about the true parameter value \( \beta \) than the undifferenced version.

Exercise 2.1. Prove (2.2) and (2.3).

Exercise 2.2. Show that \( \frac{T^{3/2}}{\mu^2}(\tilde{\beta} - \beta) \Rightarrow N(0, 3\sigma_\varepsilon^2 / \mu^2) \).